Date : 1/30/2020 10:17:28 AM
From : "Wong, Matthew C." [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org, "Lloyd, Richard E." [redacted]
Subject : [EXT] Re: nCoV analysis

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Updated again (sorry):

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – Figure 1A. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between the two at amino acid residues 350 to 550. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298) we used VirMAP to reconstruct a coronavirus genome (approximately 84% complete from samples SRR10168377 and SRR10168378) that shared 97% amino acid identity across the same RBD segment – Figure 1B. This result indicates a potential recombination event for nCoV-2019.

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Thursday, January 30, 2020 9:52 AM
To: Lloyd, Richard E. [redacted]
Cc: Wong, Matthew C. [redacted]
Subject: Re: nCoV analysis

Updated text:

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. We used VirMAP to recover potential viral genomes and compare recovered coronaviruses to the outbreak strain. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – Figure 1A. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between the two at amino acid residues 350 to 550. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298), we were able to reconstruct a coronavirus genome (approximately 84% complete from samples SRR10168377 and SRR10168378) that shared 97% amino acid identity across the same RBD segment – Figure 1B. This result indicates a potential recombination event for nCoV-2019.

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, January 30, 2020 at 9:20 AM
To: [redacted]

Cc: [redacted]
Subject: nCoV analysis

Hi Rick,
Hope you are well!

Matt and I got together last night to review his analysis on the recent nCoV-2019 genome. We came up with the following statement summarizing his findings and before posting to Virological.org we wanted to run it by you. Figures attached. Let us know what you think.

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. We used VirMAP to recover potential viral genomes and compare recovered coronaviruses to the outbreak strain. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – **Figure 1A**. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between amino acid residues 350 to 550. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298), we were able to reconstruct a coronavirus genome (approximately 84% complete from sample SRR10168377) that shared 97% amino acid identity across the same RBD genome – **Figure 1B**. This result indicates a potential recombination event for nCoV-2019.

VirMAP-Pangolin CoV genome reconstruction: *google drive link*

Best,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 1/30/2020 10:40:16 AM
From : "Lloyd, Richard E." [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org
Cc : "Wong, Matthew C." [redacted]
Subject : [EXT] Re: nCoV analysis

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi guys,
OK just got a look at this and Matt stopped by my office. I think this looks really nice and is a good way to go. You may want to include a reference for VirMAP ("VirMAP (Nature Commun. 9:3205). Go for it.
Rick

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, January 30, 2020 at 9:52 AM
To: Rick Lloyd [redacted]
Cc: "Wong, Matthew C." [redacted]
Subject: Re: nCoV analysis

Updated text:

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. We used VirMAP to recover potential viral genomes and compare recovered coronaviruses to the outbreak strain. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – Figure 1A. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between the two at amino acid residues 350 to 550. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298), we were able to reconstruct a coronavirus genome (approximately 84% complete from samples SRR10168377 and SRR10168378) that shared 97% amino acid identity across the same RBD segment – Figure 1B. This result indicates a potential recombination event for nCoV-2019.

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, January 30, 2020 at 9:20 AM
To: [redacted]
Cc: [redacted]
Subject: nCoV analysis

Hi Rick,
Hope you are well!

Matt and I got together last night to review his analysis on the recent nCoV-2019 genome. We came up with the following statement summarizing his findings and before posting to Virological.org we wanted to run it by you. Figures attached. Let us know what you think.

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. We used VirMAP to recover potential viral genomes and compare recovered coronaviruses to the outbreak strain. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – **Figure 1A**. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between amino acid residues 350 to 550. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298), we were able to reconstruct a coronavirus genome (approximately 84% complete from sample SRR10168377) that shared 97% amino acid identity across the same RBD genome – **Figure 1B**. This result indicates a potential recombination event for nCoV-2019.

VirMAP-Pangolin CoV genome reconstruction: *google drive link*

Best,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 4/16/2020 8:24:08 AM
From : "Samantha Coy" [redacted]
To : "Wilhelm, Steven W" [redacted]
Cc : "jvanetten1@unl.edu" [redacted] **, "Ajami,Nadim J"**
NAjami@mdanderson.org, [redacted] **,**
[redacted] **"**
[redacted]
"Gann, Eric" [redacted]
Subject : [EXT] Fwd: Frontiers: Congratulations! Your manuscript is accepted - 532536

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi everyone,

I think you all have received notification that our manuscript was accepted for publication, but in any case, I wanted to let everyone know as a group and pass on my gratefulness to each of you! Your contributions are much appreciated, and it feels so good to have this finished!

Hope you are all doing well with everything going on.

All the very best,

Samantha

---------- Forwarded message ---------
From: **Frontiers Microbiology Editorial Office**
<microbiology.editorial.office@frontiersin.org>
Date: Thu, Apr 16, 2020 at 4:49 AM
Subject: Frontiers: Congratulations! Your manuscript is accepted - 532536
To: <[redacted]>

Dear Dr Coy,

Frontiers Microbiology Editorial Office has sent you a message. Please click 'Reply' to send a direct response

I am pleased to inform you that your manuscript SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems has been approved for production and accepted for publication in Frontiers in Microbiology, section Virology.
Your manuscript is currently being prepared for publication. The provisional version of the abstract or introductory section is currently available online. Please do not communicate any changes at this stage. You will be contacted as soon as the author proofs are ready for your revisions.

Manuscript title: SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction

modification systems
Journal: Frontiers in Microbiology, section Virology
Article type: Original Research
Authors: Samantha R Coy, Eric Robert Gann, Spiridon E Papoulis, Michael Holder, Nadim Ajami, Joseph Petrosino, Erik Zinser, James L Van Etten, Steven W Wilhelm
Manuscript ID: 532536
Edited by: Andrew S Lang

You can click here to access the final review reports and manuscript:
http://www.frontiersin.org/Review/EnterReviewForum.aspx?activationno=80dfc921-8a82-4e86-a249-6bc5e33b7d34

As an author, it is important that you maintain your Frontiers research network (Loop) profile up to date, as your publication will be linked to your profile allowing you and your publications to be more discoverable. You can update profile pages (profile pictures, short bio, list of publications) using this link:
http://loop.frontiersin.org/people/

Tell us what you think!

At Frontiers we are constantly trying to improve our Collaborative Review process and would like to get your feedback on how we did. Please complete our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss non-profit organization:
https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?survey=author&aid=532536&uid=877766

Thank you very much for taking the time to share your thoughts.

Best regards,

Your Frontiers in Microbiology team

Frontiers | Editorial Office - Collaborative Peer Review Team
www.frontiersin.org
Avenue du Tribunal Fédéral 34, 1005 Lausanne, Switzerland
Office T 41 21 510 17 25

For technical issues, please contact our IT Helpdesk (support@frontiersin.org) or visit our Frontiers Help Center (zendesk.frontiersin.org/hc/en-us)

Date : 4/27/2020 3:35:12 PM
From : "International Journal of Gynecological Cancer" onbehalfof@manuscriptcentral.com
To : "Sims,Travis T." TTSims@mdanderson.org, "Biegert,Greyson Willis Grossman" GWBiegert@mdanderson.org, "[redacted]" "Solley,Travis N" TNSolley@mdanderson.org, "Ning,Matthew Stephen" MSNing@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Court,Kyoko" KCourt1@mdanderson.org, "Delgado Medrano,Andrea Yizel" AYDelgado@mdanderson.org, "Wu,Xiaogang" XWu10@mdanderson.org, "Ahmed-Kaddar,Mustapha" MAhmed10@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, [redacted] "Schmeler,Kathleen M" KSchmele@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Hahn,Stephen" SHahn@mdanderson.org, "Klopp,Ann H" AKlopp@mdanderson.org, [redacted]
Subject : [EXT] International Journal of Gynecological Cancer - Manuscript ID ijgc-2020-001547

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

COVID-19: A message from BMJ:
https://urldefense.com/v3/__https://authors.bmj.com/policies/covid-19__;!!PfbeBCCAmug!wyt2YFSdGRdVovVuRgT6ul2lYoVj_3wKXr_LLhQ6xZuP4Lu41Mbx6FlkN3xatsKU$

27-Apr-2020

Dear Dr. Sims:

Your manuscript entitled "Tumor Microbial Diversity and Compositional Differences Among Women in Botswana with High-Grade Cervical Dysplasia and Cervical Cancer" has been successfully submitted online and is presently being given full consideration for publication in International Journal of Gynecological Cancer.

Your manuscript ID is ijgc-2020-001547.

Please mention the above manuscript ID in all future correspondence or when calling the office for questions. If there are any changes in your street address or e-mail address, please log in to ScholarOne Manuscripts at https://urldefense.com/v3/__https://mc manuscriptcentral.com/ijgcancer__;!!PfbeBCCAmug!wyt2YFSdGRdVovVuRgT6ul2lYoVj_3wKXr_LLhQ6xZuP4Lu41Mbx6FlkN5Qigz8o$ and edit your user information as appropriate.

Please check that all author names are correctly entered as this will be the name displayed in any PubMed search.

You can also view the status of your manuscript at any time by checking your Author Center after logging in to https://urldefense.com/v3/__https://mc.manuscriptcentral.com/ijgcancer__;!!

PfbeBCCAmug!wyt2YFSdGRdVovVuRgT6ul2lYoVj_3wKXr_LLhQ6xZuP4Lu41Mbx6FlkN5Qigz8o$.

Any individuals listed as co-authors on this manuscript are copied into this submission confirmation email. If you believe that you have received this email in error, please contact the Editorial Office.

Thank you for submitting your manuscript to International Journal of Gynecological Cancer.

Respectfully,

Dr. Pedro Ramirez
Editor, International Journal of Gynecological Cancer

We are constantly trying to find ways of improving the peer review system and continually monitor processes and methods by including article submissions and reviews in our research. If you do not wish your paper or review entered into our peer review research programme, please let us know by emailing info.ijgc@bmj.com as soon as possible.

P.S. What did you think of the article submission process?
At BMJ, we constantly strive to improve our services for authors and value your feedback. We'd really like to hear your opinions as part of our on-going efforts, and we'd be grateful if could take a few minutes to fill out our short survey. Your responses will, of course, remain confidential and you won't be identified in any results.

Please click on this link to access the survey:
https://urldefense.com/v3/__https://www.surveymonkey.co.uk/r/VMXSQGP__;!!PfbeBCCAmug!wyt2YFSdGRdVovVuRgT6ul2lYoVj_3wKXr_LLhQ6xZuP4Lu41Mbx6FlkN90Gt49I$

Date : 6/17/2020 3:27:24 PM
From : "Javornik Cregeen, Sara Joan"
To : "Wong, Matthew C." , "Ajami,Nadim J"
NAjami@mdanderson.org
Cc : "Petrosino, Joseph"
Subject : [EXT] Public Virmap solution
Attachment : machineSetup.md;

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi all,

Matt and I met to discuss the status of the public Virmap set up and what still needs to be done. Matt has put together a script that will install Virmap with all dependencies on an amazon EC2 instance. I've attached a set of instructions that outline the steps to be taken and minimum requirements, etc. You might want to flesh it out a little and add any disclaimers that are needed.

What still needs doing (Matt):

- Update README.md on Virmap repo.
- Split current installer script into: basic installer, DB builder and test scripts
- Deposit scripts in github repo and add the download link to instructions (wget <path-to-installer-script>)
- Instructions for SRA tools
- Check and update "Testing the Virmap Installation"
- Do you want to include a quick note on the output files?
- Potential discrepancies in instructions:
    - Should `/scratch` be `/home/ec2-user/scratch` (that's what it is on our Amazon machine)?
    - In the "Suggested workflow" you mention creating the `TMPDIR` and setting permissions but this isn't mentioned in the instruction for the "Test run"

Thanks,
Sara

Date : 4/16/2020 9:31:21 AM
From : "James Van Etten" [redacted]
To : "Samantha Coy" [redacted]**, "Wilhelm, Steven W"** [redacted]
Cc : "Ajami,Nadim J" NAjami@mdanderson.org, " [redacted]**"** [redacted]**"Gann, Eric"** [redacted]
Subject : [EXT] Re: Frontiers: Congratulations! Your manuscript is accepted - 532536

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Great job Sam!!!!

Best personal regards,

Jim VE

From: Samantha Coy [redacted]
Date: Thursday, April 16, 2020 at 8:24 AM
To: "Wilhelm, Steven W" [redacted]
Cc: James Van Etten [redacted] Nadim Ajami <najami@mdanderson.org>, [redacted] "Gann, Eric" [redacted]
Subject: Fwd: Frontiers: Congratulations! Your manuscript is accepted - 532536

Hi everyone,

I think you all have received notification that our manuscript was accepted for publication, but in any case, I wanted to let everyone know as a group and pass on my gratefulness to each of you! Your contributions are much appreciated, and it feels so good to have this finished!

Hope you are all doing well with everything going on.

All the very best,

Samantha

---------- Forwarded message ---------
From: **Frontiers Microbiology Editorial Office** <microbiology.editorial.office@frontiersin.org>

Date: Thu, Apr 16, 2020 at 4:49 AM
Subject: Frontiers: Congratulations! Your manuscript is accepted - 532536
To: [redacted]

Dear Dr Coy,

Frontiers Microbiology Editorial Office has sent you a message. Please click 'Reply' to send a direct response

I am pleased to inform you that your manuscript SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems has been approved for production and accepted for publication in Frontiers in Microbiology, section Virology.
Your manuscript is currently being prepared for publication. The provisional version of the abstract or introductory section is currently available online. Please do not communicate any changes at this stage. You will be contacted as soon as the author proofs are ready for your revisions.

Manuscript title: SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems
Journal: Frontiers in Microbiology, section Virology
Article type: Original Research
Authors: Samantha R Coy, Eric Robert Gann, Spiridon E Papoulis, Michael Holder, Nadim Ajami, Joseph Petrosino, Erik Zinser, James L Van Etten, Steven W Wilhelm
Manuscript ID: 532536
Edited by: Andrew S Lang

You can click here to access the final review reports and manuscript:
http://www.frontiersin.org/Review/EnterReviewForum.aspx?activationno=80dfc921-8a82-4e86-a249-6bc5e33b7d34

As an author, it is important that you maintain your Frontiers research network (Loop) profile up to date, as your publication will be linked to your profile allowing you and your publications to be more discoverable. You can update profile pages (profile pictures, short bio, list of publications) using this link: http://loop.frontiersin.org/people/

Tell us what you think!

At Frontiers we are constantly trying to improve our Collaborative Review process and would like to get your feedback on how we did. Please complete our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss non-profit organization:
https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?survey=author&aid=532536&uid=877766

Thank you very much for taking the time to share your thoughts.

Best regards,

Your Frontiers in Microbiology team

Frontiers | Editorial Office - Collaborative Peer Review Team
www.frontiersin.org
Avenue du Tribunal Fédéral 34, 1005 Lausanne, Switzerland
Office T 41 21 510 17 25

For technical issues, please contact our IT Helpdesk (support@frontiersin.org) or visit our Frontiers Help Center (zendesk.frontiersin.org/hc/en-us)

Date : 4/16/2020 8:33:22 AM
From : [redacted]
To : "Samantha Coy" [redacted]**, "Wilhelm, Steven W"** [redacted]
Cc : [redacted]**"Ajami,Nadim J" NAjami@mdanderson.org, "Papoulis, Spiro"** [redacted]**"**
Subject : [EXT] Re: Frontiers: Congratulations! Your manuscript is accepted - 532536

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Congratulations, Sam!
Great news,
Erik

Erik Zinser
Associate Professor
University of Tennessee
Dept. of Microbiology
1311 Cumberland Ave
307 Ken and Blaire Mossman Bldg.
Knoxville, TN 37996-1937
Office: SERF 640
Phone: 865-974-9283
Lab: 865-974-2219
Fax: 865-974-4007
[redacted]
https://zinserlab.utk.edu

From: Samantha Coy [redacted]
Date: Thursday, April 16, 2020 at 9:24 AM
To: [redacted]
Cc: [redacted] Nadim Ajami <najami@mdanderson.org>, "Papoulis, Spiro" [redacted], "Zinser, Erik Ross" [redacted] egann [redacted]
Subject: Fwd: Frontiers: Congratulations! Your manuscript is accepted - 532536

[External Email]

Hi everyone,

I think you all have received notification that our manuscript was accepted for publication, but in any case, I wanted to let everyone know as a group and pass on my gratefulness to

each of you! Your contributions are much appreciated, and it feels so good to have this finished!

Hope you are all doing well with everything going on.

All the very best,

Samantha

---------- Forwarded message ---------
From: **Frontiers Microbiology Editorial Office**
<microbiology.editorial.office@frontiersin.org>
Date: Thu, Apr 16, 2020 at 4:49 AM
Subject: Frontiers: Congratulations! Your manuscript is accepted - 532536
To: [redacted]

Dear Dr Coy,

Frontiers Microbiology Editorial Office has sent you a message. Please click 'Reply' to send a direct response

I am pleased to inform you that your manuscript SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems has been approved for production and accepted for publication in Frontiers in Microbiology, section Virology.
Your manuscript is currently being prepared for publication. The provisional version of the abstract or introductory section is currently available online. Please do not communicate any changes at this stage. You will be contacted as soon as the author proofs are ready for your revisions.

Manuscript title: SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems
Journal: Frontiers in Microbiology, section Virology
Article type: Original Research
Authors: Samantha R Coy, Eric Robert Gann, Spiridon E Papoulis, Michael Holder, Nadim Ajami, Joseph Petrosino, Erik Zinser, James L Van Etten, Steven W Wilhelm
Manuscript ID: 532536
Edited by: Andrew S Lang

You can click here to access the final review reports and manuscript:
http://www.frontiersin.org/Review/EnterReviewForum.aspx?activationno=80dfc921-8a82-4e86-a249-6bc5e33b7d34

As an author, it is important that you maintain your Frontiers research network (Loop) profile up to date, as your publication will be linked to your profile allowing you and your publications to be more discoverable. You can update profile pages (profile pictures, short bio, list of publications) using this link: http://loop.frontiersin.org/people/

Tell us what you think!

At Frontiers we are constantly trying to improve our Collaborative Review process and would like to get your feedback on how we did. Please complete our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss non-profit organization: https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?survey=author&aid=532536&uid=877766

Thank you very much for taking the time to share your thoughts.

Best regards,

Your Frontiers in Microbiology team

Frontiers | Editorial Office - Collaborative Peer Review Team
www.frontiersin.org
Avenue du Tribunal Fédéral 34, 1005 Lausanne, Switzerland
Office T 41 21 510 17 25

For technical issues, please contact our IT Helpdesk (support@frontiersin.org) or visit our Frontiers Help Center (zendesk.frontiersin.org/hc/en-us)

Date : 4/16/2020 11:30:22 AM
From : "Papoulis, Spiro" [redacted]
To : "Samantha Coy" [redacted]
Cc : "Wilhelm, Steven W" [redacted] **, "Ajami,Nadim J" NAjami@mdanderson.org, "Erik Zinser"** [redacted] **"Gann, Eric"** [redacted]
Subject : [EXT] Re: Frontiers: Congratulations! Your manuscript is accepted - 532536

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Congratulations Sam! This is great news!

Spiridon E. Papoulis
PhD Student, Zinser Lab
Department of Microbiology
University of Tennessee - Knoxville
635 Science and Engineering Research Facility

On Thu, Apr 16, 2020 at 9:24 AM Samantha Coy [redacted] wrote:

> **[External Email]**
>
> Hi everyone,
>
> I think you all have received notification that our manuscript was accepted for publication, but in any case, I wanted to let everyone know as a group and pass on my gratefulness to each of you! Your contributions are much appreciated, and it feels so good to have this finished!
>
> Hope you are all doing well with everything going on.
>
> All the very best,
>
> Samantha
>
> ---------- Forwarded message ---------
> From: **Frontiers Microbiology Editorial Office**
> <microbiology.editorial.office@frontiersin.org>
> Date: Thu, Apr 16, 2020 at 4:49 AM
> Subject: Frontiers: Congratulations! Your manuscript is accepted - 532536
> To: [redacted]
>
> Dear Dr Coy,
>
> Frontiers Microbiology Editorial Office has sent you a message. Please click

'Reply' to send a direct response

I am pleased to inform you that your manuscript SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems has been approved for production and accepted for publication in Frontiers in Microbiology, section Virology.
Your manuscript is currently being prepared for publication. The provisional version of the abstract or introductory section is currently available online. Please do not communicate any changes at this stage. You will be contacted as soon as the author proofs are ready for your revisions.

Manuscript title: SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems
Journal: Frontiers in Microbiology, section Virology
Article type: Original Research
Authors: Samantha R Coy, Eric Robert Gann, Spiridon E Papoulis, Michael Holder, Nadim Ajami, Joseph Petrosino, Erik Zinser, James L Van Etten, Steven W Wilhelm
Manuscript ID: 532536
Edited by: Andrew S Lang

You can click here to access the final review reports and manuscript:
http://www.frontiersin.org/Review/EnterReviewForum.aspx?activationno=80dfc921-8a82-4e86-a249-6bc5e33b7d34

As an author, it is important that you maintain your Frontiers research network (Loop) profile up to date, as your publication will be linked to your profile allowing you and your publications to be more discoverable. You can update profile pages (profile pictures, short bio, list of publications) using this link:
http://loop.frontiersin.org/people/

Tell us what you think!

At Frontiers we are constantly trying to improve our Collaborative Review process and would like to get your feedback on how we did. Please complete our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss non-profit organization:
https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?survey=author&aid=532536&uid=877766

Thank you very much for taking the time to share your thoughts.

Best regards,

Your Frontiers in Microbiology team

Frontiers | Editorial Office - Collaborative Peer Review Team
www.frontiersin.org

Avenue du Tribunal Fédéral 34, 1005 Lausanne, Switzerland
Office T 41 21 510 17 25

For technical issues, please contact our IT Helpdesk (support@frontiersin.org) or visit our Frontiers Help Center (zendesk.frontiersin.org/hc/en-us)

Date : 4/16/2020 10:08:37 AM
From : "Michael E. Holder" [redacted]
To : "Samantha Coy" [redacted]
Cc : "Wilhelm, Steven W" [redacted]
"Ajami,Nadim J" NAjami@mdanderson.org, [redacted]
"Gann, Eric" [redacted]
Subject : [EXT] Re: Fwd: Frontiers: Congratulations! Your manuscript is accepted - 532536

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hello Samantha,
Congratulations!

Michael Holder

On Thu, 16 Apr 2020, Samantha Coy wrote:

> ***CAUTION:*** This email is not from a BCM Source. Only click links or open
> attachments you know are safe.
>
> _____
> Hi everyone,
> I think you all have received notification that our manuscript was accepted
> for publication, but in any case, I wanted to let everyone know as a group
> and pass on my gratefulness to each of you! Your contributions are much
> appreciated, and it feels so good to have this finished!
>
> Hope you are all doing well with everything going on.
>
> All the very best,
>
> Samantha
>
> ---------- Forwarded message ---------
> From: Frontiers Microbiology Editorial Office
> <microbiology.editorial.office@frontiersin.org>
> Date: Thu, Apr 16, 2020 at 4:49 AM
> Subject: Frontiers: Congratulations! Your manuscript is accepted - 532536
> To: [redacted]
>
>
> Dear Dr Coy,
>
> Frontiers Microbiology Editorial Office has sent you a message. Please click
> 'Reply' to send a direct response
>
> I am pleased to inform you that your manuscript SMRT sequencing of
> Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability
> in adenines targeted by restriction modification systems has been approved
> for production and accepted for publication in Frontiers in Microbiology,
> section Virology.
> Your manuscript is currently being prepared for publication. The provisional
> version of the abstract or introductory section is currently available

> online. Please do not communicate any changes at this stage. You will be
> contacted as soon as the author proofs are ready for your revisions.
>
> Manuscript title: SMRT sequencing of Paramecium bursaria Chlorella Virus-1
> reveals diverse methylation stability in adenines targeted by restriction
> modification systems
> Journal: Frontiers in Microbiology, section Virology
> Article type: Original Research
> Authors: Samantha R Coy, Eric Robert Gann, Spiridon E Papoulis, Michael
> Holder, Nadim Ajami, Joseph Petrosino, Erik Zinser, James L Van Etten,
> Steven W Wilhelm
> Manuscript ID: 532536
> Edited by: Andrew S Lang
>
> You can click here to access the final review reports and
manuscript:https://urldefense.com/v3/__http://www.frontiersin.org/Review/EnterReviewForum.aspx?
activationno=80dfc9__;!!PfbeBCCAmug!19teF0QhFLruG-
Pq9sFLvqlyJfEr7CPsjuccPq2po7_OTuf9MFYiuVqFEUTeD1np$
> 21-8a82-4e86-a249-6bc5e33b7d34
>
> As an author, it is important that you maintain your Frontiers research
> network (Loop) profile up to date, as your publication will be linked to
> your profile allowing you and your publications to be more discoverable. You
> can update profile pages (profile pictures, short bio, list of publications)
> using this link: https://urldefense.com/v3/__http://loop frontiersin.org/people/__;!!PfbeBCCAmug!
19teF0QhFLruG-Pq9sFLvqlyJfEr7CPsjuccPq2po7_OTuf9MFYiuVqFEd2UkX2r$
>
> ----------
> Tell us what you think!
>
> At Frontiers we are constantly trying to improve our Collaborative Review
> process and would like to get your feedback on how we did. Please complete
> our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss
> non-profit organization:
> https://urldefense.com/v3/__https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?
survey=author&a__;!!PfbeBCCAmug!19teF0QhFLruG-
Pq9sFLvqlyJfEr7CPsjuccPq2po7_OTuf9MFYiuVqFEZqp1m59$
> id=532536&uid=877766
>
> Thank you very much for taking the time to share your thoughts.
>
> Best regards,
>
> Your Frontiers in Microbiology team
>
> Frontiers | Editorial Office - Collaborative Peer Review Team
> https://urldefense.com/v3/__http://www.frontiersin.org__;!!PfbeBCCAmug!19teF0QhFLruG-
Pq9sFLvqlyJfEr7CPsjuccPq2po7_OTuf9MFYiuVqFEXq8tRgz$
> Avenue du Tribunal Fédéral 34, 1005 Lausanne, Switzerland
> Office T 41 21 510 17 25
>
> For technical issues, please contact our IT Helpdesk
> (support@frontiersin.org) or visit our Frontiers Help Center
> (zendesk frontiersin.org/hc/en-us)
>
>

Date : 4/16/2020 6:57:33 AM
From : "Hoffman, Kristi Louise" [redacted]
To : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org
Cc : "Wong, Matthew C." [redacted] **Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org**
Subject : [EXT] Re: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

This request is in Sara's queue, and she will complete it as soon as her urgent COVID tasks are done. She expects to have it Friday.

Kristi

From: "Khan,Md Abdul Wadud" <MKhan7@mdanderson.org>
Date: Saturday, April 11, 2020 at 9:19 PM
To: "Hoffman, Kristi Louise" [redacted]>
Cc: "Wong, Matthew C." [redacted]>, "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wargo,Jennifer" <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

Hope you are staying safe and healthy.

Wondering whether you have any update on the metaphlan2?

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Friday, March 27, 2020 10:52 AM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted] Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>; Petrosino, Joseph [redacted]
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud (and team),

The earliest the MetaPhlAn2 request can be completed is the week of April 6th. Let me know if you'd still like us to process the data given that timeframe.

Please note that with regards to Virmap, data processing requests need to go through a project manager and completed according to our queue. While we can expedite requests, especially for *trusted*, long-term collaborators, proper procedures still need to be followed. Circumventing these procedures affects other valued CMMR collaborators and is not taken lightly. I expect this won't be an issue going forward and any requests will go through the proper channels.

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 3:44 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am actually hoping to get the output of MetaPhlan2 by this week but if you can get it done by next week that would be great too.

I already got the output of VirMap. So, no worry on this analysis.

Best

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Wednesday, March 25, 2020 2:47 PM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted]>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run them through the MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-

mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/18/2020 4:45:05 PM
From : "Javornik Cregeen, Sara Joan" [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org, "Petrosino, Joseph" [redacted]**, "Hoffman, Kristi Louise"** [redacted]**,**
"Wong, Matthew C." [redacted]
Subject : [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/26/2020 1:52:16 PM
From : "Javornik Cregeen, Sara Joan" [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org, "Hoffman, Kristi Louise" [redacted] **"Petrosino, Joseph"** [redacted]
Subject : [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]

Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you

are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 2/13/2020 7:57:16 AM
From : "Ajami,Nadim J"
To : [redacted] **"Joseph Petrosino"** [redacted]
"Matthew C. Wong" [redacted]
Subject : Fwd: [EXT] bioRxiv -- Manuscript Closed

Sent from my iPhone

Begin forwarded message:

From: [redacted]
Date: February 13, 2020 at 12:09:50 AM CST
To: "Ajami,Nadim J" <NAjami@mdanderson.org>
Subject: [EXT] bioRxiv -- Manuscript Closed

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

MS ID#: BIORXIV/2020/925941
MS TITLE: nCoV Spike Protein S1 CTD subdomain Shares High Amino Acid Identity With a Coronavirus Recovered from a Pangolin Viral Metagenomic Dataset

Dear Nadim Ajami;

The above manuscript has been closed.

The bioRxiv team

Date : 4/15/2020 7:20:48 PM
From : "Sims,Travis T." TTSims@mdanderson.org
To : "Sastry,Jagannadha K" jsastry@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Schmeler,Kathleen M" KSchmele@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "Chapman,Bhavana S" BSChapman@mdanderson.org, "Sastry,Jagannadha K" jsastry@mdanderson.org, "Mezzari, Melissa" [redacted] "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Schmeler,Kathleen M" KSchmele@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, [redacted]
Cc : "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org
Subject : Manuscript - Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation
Attachment : Manuscript - Gut microbiome diversity as an independent predictor of survival in cervical cancer patients receiving chemoradiation V1.docx;Table 1. Gut Microbiome Univariate and Multivariate Analysis RFS 4-15-20.docx;Table 2. Gut Microbiome Univariate and Multivariate Analysis OS 4-15-20.docx;Figures V1 - Gut microbiome diversity an independent predictor of cervical cancer 4-15-2020.pptx;Supplemental Table 1. Baseline diversity vs. demographics 4-15-20.docx;Supplemental Table 2. Gut Microbiome 4-15-20.docx;Supplemental Table 3. Gut Microbiome Univariate All Alpha Diversity Time Points RFS 4-15-20.docx;Supplemental Table 4. Gut Microbiome Univariate All Alpha Diversity Time Points OS 4-15-20.docx;

Hello all,
We have completed the first draft of the manuscript for our project “Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation”.

Mrs. El Alam, Dr. Colbert, Dr. Klopp and I have included you on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to *Nature Medicine*.

Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.

I am grateful for your time and feedback regarding this project! We hope to submit by 5/1/20.
Best,
Travis

Travis T. Sims, MD, MPH

Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

TITLE: Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation

Authors and Affiliations

Travis T. Sims, MD, MPH[1*], Molly B. El Alam, MPH[2*], Tatiana V. Karpinets, PhD[3], Kyoko Yoshida-Court, PhD[2], Xiaogang Wu, PhD[3], Greyson W. G. Biegert, BS[2], Andrea Y. Delgado Medrano, BS[2], Travis Solley, BS[2], Mustapha Ahmed-Kaddar, BS[2], Bhavana V. Chapman, MD[2], Jagannadha Sastry, PhD[4], Melissa P. Mezzari, PhD[5], Joseph F. Petrosino, PhD[5], Lilie L. Lin, MD[2], Lois Ramondetta, MD[1], Anuja Jhingran, MD[2], Kathleen Schmeler, MD[1], Nadim J Ajami, PhD[3], Jennifer Wargo, MD, MMSc[6], Lauren E. Colbert, MD, MSCR[2+,] Ann H. Klopp, MD, PhD[2+]

[1]Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [2]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [3]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [4]Department of Immunology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [5]Department of Molecular Virology and Microbiology, Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA. [6]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

* Authors Contributed Equally

+Shared corresponding authorship

Correspondence: L.E. Colbert or A.H. Klopp, Department of Radiation Oncology, Unit

1422, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA. Telephone: 832-652-6033 (L.E.C.), 713-563-2444 (A.H.K); fax: 713-745-2398; e-mail: lcolbert@mdanderson.org, aklopp@mdanderson.org.

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Research Support

This research was supported in part by the National Institutes of Health (NIH) through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant 5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK).

Role of Funding Sources

The funding sources were not involved in the research hypothesis development, study design, data analysis, or manuscript writing. Data access was limited to the authors of this manuscript.

ABSTRACT

Background: Diversity of the gut microbiome is associated with response rates for patients with melanoma receiving immunotherapy and chemotherapy but has not been investigated in patients receiving radiation therapy. Additionally, studies investigating the gut microbiome and outcomes in cancer patients may not adjusted for established risk factors. We sought to determine if diversity and composition was independently associated with survival in cervical cancer (CC) patients receiving chemoradiation (CRT).

Methods: We analyzed baseline 16S rDNA fecal microbiomes of CC patients receiving standard CRT. Cervical tumor brushings were analyzed using flow cytometry. Patient and tumor characteristics were analyzed by univariate and multivariate Cox regression models for recurrence-free survival (RFS) and overall survival (OS) based on univariate p-value < 0.2. Characteristics included age, body mass index (BMI), race, stage, grade, histology, nodal status, and max tumor size. Alpha (within sample) diversity was evaluated using Shannon diversity index (SDI). Kaplan-Meier curves were generated for patients with high and normal BMI and overweight/obese BMI based on Cox analysis.

Results: 55 CC patients were included. Univariate analysis identified older age (Hazard Ratio (HR) of 0.93 (95% CI = 0.87-0.98, $P = 0.0096$)), SDI (HR of 0.51 (95% CI = 0.23-1.1, $P = 0.087$)) and BMI (HR of 0.92 (95% CI = 0.84-1, $P = 0.096$)) as risk factors for RFS. Multivariate survival analyses identified BMI and SDI as independent prognostic factors for RFS with a HR of 0.87 (95% CI = 0.77-0.98, $P = 0.02$) and 0.36 (95% CI = 0.15-0.84, $P = 0.018$) respectively. For OS,

multivariate survival analyses again identified BMI and SDI as independent prognostic factors with a HR of 0.78 (95% CI = 0.623-0.97, *P* = 0.025) and 0.19 (95% CI = 0.043-0.83, *P* = 0.028) For all patients, multiple taxa differed markedly between short term and long term survivors. Short term survivor fecal samples were significantly enriched in *porphyromonas, porphyromonadaceae, and dialister*, whereas long term survivor samples were significantly enriched in *Escherichia Shigella, Enterobacteriaceae, and Enterobacteriales* (P < 0.05; LDA score > 3.5) Analysis of cervical tumor brush flow cytometry revealed that patients with a high microbiome diversity had increased infiltration of CD4+ lymphocytes and well as activated subsets of CD4 cells expressing ki67+ and CD69+ over the course of radiation therapy.

Conclusion: Gut diversity is a significant predictor of OS in CC patients undergoing CRT and compositional differences were observed between patients who were short and long term survivors. Patients with high gut microbial diversity exhibit enhanced T cell signatures. Studies are needed to determine if modification of the gut microbiome will improve outcomes for women with cervical cancers.

Key words: gynecologic cancer, microbiome, chemoradiation

INTRODUCTION

Cervical cancer continues to be one of the leading causes of cancer-associated mortality globally[1]. In the United States, more than 13,000 women will be diagnosed with invasive cervical cancer in 2019, resulting in more than 4,250 deaths[2]. Multimodality therapy consisting of concurrent chemoradiation (CRT) comprising external-beam radiotherapy (EBRT) and systemic chemotherapy followed by intracavitary brachytherapy continues to be the standard of care in clinical practice for locally advanced disease[3].

The fecal or gut microbiome, a diverse community of bacteria, archaea, fungi, protozoa, and viruses, is thought to influence host immunity by modulating multiple immunologic pathways, thus impacting health and disease[4–6]. Studies have suggested that dysbiosis of the gut microbiome confers a predisposition to certain malignancies and influences the body's response to a variety of cancer therapies, including chemotherapy, radiotherapy, and immunotherapy[6–10]. For example, melanoma patients are more likely to have a favorable response to immune checkpoint blockade and exhibit improved systemic and antitumor immunity if they have a more diverse intestinal microbiome[10].

Radiotherapy promotes the activation of T cells directed against tumor antigens[11–14]. In combination with immunotherapy, radiotherapy can maximize the antitumor immune response and promote durable disease control[15,16]. We theorize that the gut microbiota may modulate radioresponse through immunologic mechanisms[13,17]. Studies investigating the gut microbiome and outcomes in cancer patients often do not adjust for confounding patient and tumor characteristics. To assess this, we sought to identify independent gut microbial risk factors in cervical cancer (CC) patients receiving chemoradiation (CRT) and to evaluate their impact on

survival. We hypothesize that gut microbial differences may affect clinical outcomes in patients with cervical cancer.

RESULTS

Patient Characteristics

A total of 55 patients with a mean age of 47 years (range, 29-72 years) volunteered to participate in this study. The patients received standard treatment for cervical cancer with 5 weeks of EBRT and weekly cisplatin. After completion of EBRT, patients received brachytherapy. For evaluation of treatment response, patients underwent magnetic resonance imaging (MRI) at baseline and week 5 and positron emission tomography (PET)/computed tomography (CT) 3 months after treatment completion (Fig. 1a). Most patients had stage IIB disease (51%) and squamous histology (78%). Their clinicopathologic data are summarized in Supplementary Table 1. We staged cervical cancer using the 2014 International Federation of Gynecology and Obstetrics staging system. The median cervical tumor size according to MRI was 5.4 cm (range, 1.2-11.5 cm). Thirty patients (55%) had lymph node involvement according to PET or CT. We first analyzed the bacterial 16S rDNA (16Sv4) fecal microbiota at baseline with respect to disease histology, grade, and stage. We found that the baseline α-diversity (within tumor samples) and β-diversity (between samples) of the fecal microbiome in the cervical cancer patients did not differ according to histology, grade, or stage ($P > 0.05$) (Supplementary Fig. 1a-d).

Univariate and multivariate analysis of factors affecting recurrence free survival (RFS) and overall survival (OS)

In the univariate Cox proportional hazard regression model predicting RFS, 3 covariates showed $p \leq 0.2$. As shown in Table I, univariate analysis identified older age (Hazard Ratio (HR) of 0.93 (95% CI = 0.87-0.98, P = 0.0096)), SDI (HR of 0.51 (95% CI = 0.23-1.1, P = 0.087)) and BMI (HR of 0.92 (95% CI = 0.84-1, P = 0.096)) as risk factors for RFS. Multivariate survival analyses identified BMI and SDI as independent prognostic factors for RFS with a HR of 0.87 (95% CI = 0.77-0.98, P = 0.02) and 0.36 (95% CI = 0.15-0.84, P = 0.018) respectively. As shown in Table 2, univariate analysis identified SDI (HR of 0.34 (95% CI = 0.1-1.1, P = 0.08) and BMI (HR of 0.83 (95% CI = 0.69-1, P = 0.055)) as risk factors for OS. For OS, multivariate survival analyses again identified BMI and SDI as independent prognostic factors with a HR of 0.78 (95% CI = 0.623-0.97, P = 0.025) and 0.19 (95% CI = 0.043-0.83, P = 0.028) respectively.

Baseline Gut Microbiota Diversity is Associated with Favorable Responses

During the median follow-up period of 24.5 months, 7 patients died; all patients (12.7% of the total study population) died of disease (DOD). Figure 1 shows the Kaplan-Meier curves for RFS and OS. Given that in our univariate and multivariate analyses performed by Cox proportional hazard model Shannon index was confirmed as an independent predictor for RFS and OS, we first tested the relationship between diversity and RFS and OS in our cohort by stratifying patients based on high and low Shannon diversity metric. We stratified the patients by Shannon index as high-diversity versus low-diversity groups based on the cutoff value of Shannon index (2.69) calculated by receiver operating characteristic curve (ROC). We demonstrate that patients with high fecal alpha diversity at baseline showed a trend toward prolonged RFS and OS when compared to those with low diversity (P = 0.16 and 0.094, respectively) (Fig 1a,b). Next, because our univariate and multivariate analyses performed by Cox proportional hazard model also identified BMI as an independent predictor for RFS and OS we tested the relationship between

diversity and RFS and OS in our cohort by stratifying patients based on high and low Shannon diversity metric and normal or high BMI. As shown in Figure 1d,e, when BMI and gut diversity are stratified for at baseline, patients with normal BMI and higher SDI had a longer median RFS duration (P = 0.0027) (Fig 1d). OS (Fig 1e). Overall survival was longer for patients with normal BMI and higher gut diversity (P = 0.2).

Compositional Difference in Gut Microbiome in Response to chemoradiation

To further investigate whether the composition of gut microbiome was associated with the response to CRT, we used Linear discriminant analysis (LDA) Effect Size analysis to identify bacterial genera that were differentially enriched in short term and long term cervical cancer patients ($P < 0.05$; LDA score > 3.5). In all patients, multiple taxa differed significantly at baseline between short and long term survivors. Specifically, short term survivor fecal samples were significantly enriched in *porphyromonas*, *porphyromonadaceae, and dialister*, whereas long term survivor samples were significantly enriched in *Escherichia Shigella, Enterobacteriaceae, and Enterobacteriales* ($P < 0.05$; LDA score > 3.5, Fig 2a,b). Given that in our univariate analyses performed by Cox proportional hazard model *Pasteurellales, Haemophilus and Veillonella* were confirmed as an independent predictor for RFS and OS, we tested the relationship between these taxa and RFS and OS in our cohort by stratifying patients based on their relative abundance at baseline (Supplemental Fig 2). We demonstrate that patients with high relative abundance of *Veillonella* at baseline showed a trend toward prolonged RFS and OS when compared to those with a low relative abundance at baseline (P = 0.08 and P = 0.054, respectively).

Association between Gut Microbiota Profile and Immune Signatures

Because the gut microbiota is thought to influence disease progression partially through modulating systemic immune responses, we analyzed the cervical tumors in our cohort of patients via flow cytometry on tumor brushings performed before week 1, week 3 and week 5 of radiation therapy. To identify features associated with high gut diversity, Spearman correlation analysis was conducted between immune signatures at each time point. High Shannon diversity index was positively correlated with tumor infiltration of CD4 T cells at week 3, CD4ki67+ T-cells at week 5, (Table 3 and Fig 4a-d). The results suggest that patients with high gut diversity develop increased infiltration of activated CD4+ T-cell subsets.

DISCUSSION

The aim of this study was to identify independent gut microbial risk factors in cervical cancer patients receiving chemoradiation and to evaluate their impact on survival. We found BMI and gut diversity to be independent risk factors for RFS and OS in cervical cancer patients undergoing chemoradiation. The results indicate that overweight or obesity is a favorable prognostic factor independent of gut diversity. Additionally, our results demonstrate that patients with better clinical survival exhibit higher diversity as well as a distinct gut microbiome composition. Lastly the association between gut microbiome diversity and systematic immune signatures highlights helper CD4+ T cells as potential mediators of antitumor immunity upon CRT treatment.

Authors have previously described the gut microbiome and its effect on treatment outcomes for a variety of malignancies[39–41]. The diversity of gut microbiome is defined as the number and abundance distribution of distinct types of microorganisms colonizing within the gut[18]. In our study, higher alpha diversity at baseline correlated with an improved RFS and OS. High diversity implies more species harbor in the gut and suggests a difference in gut composition between short term and long term survivors. Our results imply that the diversity of gut microbiota

might be a shared benefit factor in those who respond well to CRT treatment. It is now generally accepted that the gut microbiome modulates immune responses, antitumor immunity, and clinical outcomes in a variety of malignancies[8,10,19]. The gut microbiome is thought to affect both innate and adaptive immune responses. Specifically how the gut microbiome exerts its influence continues to be explored, but this explanation may have important implications if specific taxa are found to change host response to treatment via immunomodulation[6]. In our study, T helper cell profiles at baseline correlate with gut diversity. These results confer that T cells and response to CRT are likely affected by the gut microbiota independent of other factors such as BMI. Using multi-color flow cytometry we performed correlation analysis on individual immune signatures and microbiota diversity. The frequency of helper CD4+ T cells were chiefly identified. Cervical cancer is considered to be an immunogenic tumor because its origin is dependent on a persistent infection with human papilloma virus (HPV), most often HPV16 or HPV18[20]. Previous studies have reported that the number and functional orientation of tumor-infiltrating CD4+ and CD8+ T cells and the presence of M1 type macrophages strongly correlates with survival in patients with cervical cancer after chemoradiation[20,21]. T cells are capable of rapid antigen-specific responses and play critical roles in immune recall responses. In addition to the percentage of CD4+ t cell subsets, the increase in CD4 Ki67, CD4 CD69, and CD4 PD1 in the patients with high microbiota diversity implies that gut microbiome also modulates the proliferation of certain immune cell populations. Recent studies have already reported that chemoradiotherapy for cervical cancer induces unfavorable immune changes reflected by a decreased number of circulating lymphocytes, both CD4+ and CD8+ T cells, and an increased percentage in myeloid-cell populations, including myeloid-derived suppressor cells and monocytes[20]. Whereas CD4+ T cells infiltrating in tumor microenvironment are thought to help the activity of other immune cells by releasing T cell

cytokines, circulating CD4+ T cell subsets reported here are probably inclined to reflect the role of gut microbiota on systemic immune responses. How peripheral memory CD4+ T cell signatures affect the efficacy of CRT treatment needs to be investigated in the future. Our study shows that the diversity of gut microbiota is associated with favorable response to CRT against cervical cancer. Considering the correlation between microbiota diversity and peripheral helper T cells being reshaped upon CRT treatment, we propose that patients with more diverse gut microbiota at baseline may benefit from CRT to a greater extent. This might be mediated by reprogramming systemic antitumor immune responses. The significance of our study lies in that the modulation of gut microbiota before treatment might provide an alternative way to enhance the efficacy of CRT, specifically in cases with positive lymph nodes and advanced stages in which systemic failure of current therapies represents a major challenge. Our results suggest that changes in the gut microenvironment contribute substantially to treatment success or failure, particularly in so-called immunogenic tumors like cervical cancer. Additionally, there is emerging data describing the influence of the gut microbiome as it pertains to radiotherapy[22]. Given that radiation can change the composition of the gut microbiome by altering the relative abundance of different taxa, we have to postulate whether it is these changes that ultimately alter the effectiveness of radiotherapy for cervical cancer[6,23,24].

In our cohort, at baseline, a higher relative abundance of *Veillonella* resulted in a trend toward prolonged RFS and OS. Our own group has previously characterized the 16S rDNA fecal microbiome cervical cancer patients compared to healthy female controls, and have reported on differences in the relative abundance of specific taxa[42]. Our new findings support the hypothesis that organisms like *Veillonella* inhabiting the gut microbiome may be manipulated to improve cancer treatment response. Knowing specific gut microbial organisms that inhabit and undergo

changes in patients with cervical cancer during CRT provides further insight into mechanisms that may modulate immune response and potentiate treatment outcomes in cancer patients. The results of our study illustrate the potential of intentionally modifying the gut microbiota to accumulate CRT-tolerant species as an interventional strategy to enhance response of cervical cancer to CRT. Researchers have studied the treatment-enhancing utility of the gut microbiota in multiple areas of medicine[9,38]. For example, human fecal microbial transplants have protected germ-free mice from arsenic-induced mortality and reduced the number of antibiotic-resistant genes in patients with recurrent *Clostridium difficile* infections[39,40]. Also, Wang et al.[41] recently reported on the first case series of patients with immune checkpoint inhibitor-associated colitis successfully treated with fecal microbiota transplantation. With respect to how the gut microbiome can modulate the host response to chemotherapy, a previous review highlighted three important clinical elements: facilitation of drug efficacy, compromise of anticancer effects, and mediation of toxicity[42]. The authors went on to predict how the gut microbiome could be modified in clinical practice to increase cancer treatment efficacy and reduce toxicity. For example, in a murine model, radiation-induced dysbiosis increased the susceptibility of mice to radiotherapy-related gastrointestinal toxic effects[23]. Determining whether changes in the human gut microbiome during CRT affect patients' risk of treatment-related toxic effects may be an area for further investigation.

The "obesity paradox", which suggest a positive effects of increasing BMI as it pertains to a specific disease, was firstly reported in heart failure[25], but has since been described in a variety of disease processes including coronary artery disease, kidney disease, diabetes, and a variety of malignancies, including other gynecologic cancers[26–28]. Theories centered around the "obesity paradox" suggest that patients with a high BMI may be better able to withstand cancer-induced consumption and stress compared with patients with a low BMI[37]. Other theories include greater

metabolic reserve, an attenuated response to hormones involved in the renin–angiotensin–aldosterone system, fitness and its association with adiposity and clinical prognosis, and unmeasured confounding factors[26]. For example, in uterine cancer it has been reported that the risk of recurrence differed significantly by BMI[29]. Specifically, a greater proportion of obese women (BMI ≥ 40) met criteria for having a low risk of recurrence, while thin women tended to have a high intermediate risk or recurrence. There have been many studies investigating the impact of BMI on cervical cancer, but the association between BMI and cervical cancer remains unclear[30]. Most cervical cancer is caused by a persistent infection with a high risk human papillomavirus (HPV). However, it has been suggested that obesity may increase the risk of cervical cancer[31]. Contributing factors include poor screening and that body fat distribution hormonally influences the risk of glandular cervical carcinoma like adenocarcinoma of the cervix[32,33].

In contrast, Brinton et al reported that body weight was not an independent prognostic factor for squamous cell tumors, and a slight increased risk of adenocarcinoma, although this was not significant[34]. Tornberg et al. reported that there was not a significant relationship between being overweight and cervical cancer[35] and a review conducted in 2008 by Lane et al. did not report a relationship between cervical cancer and obesity siting a of a lack of evidence[36]. Finally, a meta-analysis done by Poorolajal et al. in 2016 indicated that being overweight (BMI 25–29.9 kg/m2), is not associated with an increased risk of cervical cancer, but that obesity (BMI ≥30 kg/m2) is weakly associated with an increased risk of cervical cancer[30]. However, the authors warned that more evidence, based on large prospective cohort studies, is required to provide conclusive evidence on whether or not BMI is associated with an increased risk of cervical cancer. These factors demonstrate the need to better understand if and how obesity increases cervical cancer risk. The inconsistent conclusions among studies investigating the association between

BMI and cervical cancer may be attributed to numerous factors including, but not limited to, patient selection criteria, sample size and generalizability of the study population to the general public. Among these factors, patient selection criteria may be especially important, because tumor histology seems to be closely associated with BMI.

The strengths of this study include the use of careful clinical staging, histopathology, and reliable phylogenetic and statistical analysis to assess bacterial community compositional changes using both microbial divergence and taxon-based methods. Additionally, we followed a complete protocol for 16S sequencing ranging from the sample collection method to DNA extraction and sequencing, thus limiting artifactual variations. Although this study yielded intriguing findings, it was limited by its small sample. Consequently, the sample size limited our ability to weigh statistical power. However, results presented herein provide solid evidence of the effect of CRT on the gut microbiome.

In conclusion, our study demonstrated that gut diversity is a significant factor for predicting OS in CC patients undergoing CRT when BMI is accounted for, and may help explain the "obesity paradox" in cancer response. Our study shows that the diversity of gut microbiota is associated with a favorable response to chemoradiation against cervical cancer. Considering the correlation between microbiota diversity and T cells being influenced with CRT treatment, patients with more diverse gut microbiota at baseline may benefit from CRT to a greater extent. The significance of our study lies in that the modulation of gut microbiota before CRT might provide an alternative way to enhance the efficacy of CRT but this needs to be validated in large cohort studies. Studies exploring the relationship between gut diversity, CRT, and treatment efficacy are needed to further understand the role of the gut microbiome in treatment outcomes.

ONLINE METHODS

Participants and clinical data. Gut microbiome and cervical swab samples were collected prospectively from cervical cancer patients according to a protocol approved by The University of Texas MD Anderson Cancer Center Institutional Review Board (MDACC 2014-0543) for patients with biopsy-proven carcinoma of the cervix treated at MD Anderson and the Lyndon B. Johnson Hospital Oncology Clinic from September 22, 2015, to January 11, 2019. All patients had new diagnoses of locally advanced, nonmetastatic carcinoma of the cervix and underwent definitive CRT with EBRT followed by brachytherapy. Patients received a minimum of 45 Gy via EBRT in 25 fractions over 5 weeks with weekly cisplatin followed by two brachytherapy sessions at approximately weeks 5 and 7 with EBRT in between for gross nodal disease or persistent disease in the parametrium. Patients with stage IB1 cancer were given CRT due to the presence of nodal disease. Clinical variables, demographics, and pathologic reports were abstracted from electronic medical records.

Sample collection and DNA extraction. Stool was collected from all patients by a clinician performing rectal exams at five time points (baseline; weeks 1, 3, and 5 of radiotherapy; and 3 months after CRT completion) using a matrix-designed quick-release Isohelix swab to characterize the diversity and composition of the microbiome over time. The swabs were stored in 20 µl of protease K and 400 µl of lysis buffer (Isohelix) and kept at -80°C within 1 h of sample collection.

16S rRNA gene sequencing and sequence data processing. 16S rRNA sequencing was performed for fecal samples obtained from all patients at four time points to characterize the diversity and composition of the microbiome over time. 16S rRNA gene sequencing was done at

the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine. 16S rRNA was sequenced using approaches adapted from those used for the Human Microbiome Project[43]. The 16S rDNA V4 region was amplified via polymerase chain reaction with primers that contained sequencing adapters and single-end barcodes, allowing for pooling and direct sequencing of polymerase chain reaction products. Amplicons were sequenced on the MiSeq platform (Illumina) using the 2 x 250-bp paired-end protocol, yielding paired-end reads that overlapped nearly completely. Sequence reads were demultiplexed, quality-filtered, and subsequently merged using the USEARCH sequence analysis tool (version 7.0.1090) (4). 16S rRNA gene sequences were bundled into operational taxonomic units at a similarity cutoff value of 97% using the UPARSE algorithm[44]. To generate taxonomies, operational taxonomic units were mapped to an enhanced version of the SILVA rRNA database containing the 16Sv4 region. A custom script was used to create an operational taxonomic unit table from the output files generated as described above for downstream analyses of α-diversity, β-diversity, and phylogenetic trends. Principal coordinates analysis was performed by institution and sample set to make certain no batch effects were present.

Flow Cytometry. Immunostaining was performed according to standard protocols. Cells were fixed using the Foxp3/Transcription Factor Staining Buffer Set (eBioscience) and stained with a 16 color panel with antibodies from Biolegend, BD Bioscience, eBioscience, and Life Technologies. Analysis was performed on a 5-laser, 18 color LSRFortessa X-20 Flow Cytometer (BD Biosciences). Analysis was performed on Flowjo Software (INFO). Briefly, cells were stained with intracellular mAB for 30 minutes at 4C in the presence of anti-Cd16/Cd32 mAB (BD Bioscience), fixed with Foxp3/Transcription Factor Staining Buffer Set (eBioscience), and held in

FACS (Corning, 2 mM EDTA, 2% FBS). Counting beads (Thermo Fisher) were used for single color controls.

Statistical analyses. For microbiome analysis, rarefaction depth was set at 7066 reads. The ISD index was used to evaluate α-diversity (within samples), and principle coordinates analysis of unweighted UniFrac distances was used to examine β-diversity (between samples). Patient and tumor characteristics were analyzed by univariate and multivariate Cox regression models for Recurrence-free survival (RFS) and Overall survival (OS) based on univariate p-value < 0.2. Characteristics included age, body mass index (BMI), race, stage, grade, histology, nodal status, smoking status, antibiotic use and max tumor size. For each outcome of interest, a multivariate Cox regression analysis was performed to adjust for the effects of prognostic factors identified on univariate analysis as influencing survival in cervical cancer. These analyses were conducted using covariates with $p \leq 0.2$ in a stepwise fashion. Alpha (within sample) diversity was evaluated using Shannon diversity index (SDI). The relative abundance of microbial taxa, classes, and genera was determined using LDA Effect Size[45], applying the one-against-all strategy with a threshold of 2 for the logarithmic LDA score for discriminative features and α of 0.05 for factorial Kruskal-Wallis testing among classes. LDA Effect Size analysis was restricted to bacteria present in 20% or more of the study population. Kaplan-Meier curves were generated for patients with normal BMI and overweight/obese BMI based on Cox analysis and clostridia abundance. The significance of differences was determined using the log-rank test. Gut microbial diversity, RFS, and OS were also compared using Pearson correlation, linear regression, and Cox regression analysis.

ACKNOWLEDGEMENTS

This work was supported in part by the Radiological Society of North America Resident/Fellow Award (to L.E.C.), the NIH/NCI under award number P30CA016672, and an NIH T32 grant (#5T32 CA101642-14; to T.T.S.). This study was partially funded by the MD Anderson HPV-Related Cancers Moon Shot program (to L.E.C. and A.K.). The human subjects who participated in this study are gratefully acknowledged.

AUTHOR CONTRIBUTION

All authors were involved in subject identification, data collection, interpretation of the statistical analysis, and review and approval of the final manuscript. The study concept was conceived by L.E.C., A.K., and T.T.S. The manuscript was written by T.T.S.

COMPETING INTERESTS

The authors report no conflicts of interest, financial or otherwise, related to the subject of this article.

ROLE OF FUNDING SOURCES

The funding sources were not involved in the development of the research hypothesis, study design, data analysis, or writing of the manuscript. Data access was limited to the authors of the manuscript.

References:

1. Islami, F., Fedewa, S. A. & Jemal, A. Trends in cervical cancer incidence rates by age, race/ethnicity, histological subtype, and stage at diagnosis in the United States. *Prev. Med.* **123**, 316–323 (2019).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* **69**, 7–34 (2019).
3. Monk, B. J., Tewari, K. S. & Koh, W.-J. Multimodality therapy for locally advanced cervical carcinoma: state of the art and future directions. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25**, 2952–2965 (2007).
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
5. Cho, J. A. & Chinnapen, D. J. F. Targeting friend and foe: Emerging therapeutics in the age of gut microbiome and disease. *J. Microbiol.* **56**, 183–188 (2018).
6. Helmink, B. A., Khan, M. A. W., Hermann, A., Gopalakrishnan, V. & Wargo, J. A. The microbiome, cancer, and cancer therapy. *Nat. Med.* **25**, 377–388 (2019).
7. Cerf-Bensussan, N. & Gaboriau-Routhiau, V. The immune system and the gut microbiota: friends or foes? *Nat. Rev. Immunol.* **10**, 735–744 (2010).
8. McQuade, J. L., Daniel, C. R., Helmink, B. A. & Wargo, J. A. Modulating the microbiome to improve therapeutic response in cancer. *Lancet Oncol.* **20**, e77–e91 (2019).
9. Roy, S. & Trinchieri, G. Microbiota: a key orchestrator of cancer therapy. *Nat. Rev. Cancer* **17**, 271–285 (2017).
10. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018).
11. Dewan, M. Z. *et al.* Fractionated but Not Single-Dose Radiotherapy Induces an Immune-Mediated Abscopal Effect when Combined with Anti-CTLA-4 Antibody. *Clin. Cancer Res.* **15**, 5379–5388 (2009).
12. Lugade, A. A. *et al.* Local Radiation Therapy of B16 Melanoma Tumors Increases the Generation of Tumor Antigen-Specific Effector Cells That Traffic to the Tumor. *J. Immunol.* **174**, 7516–7523 (2005).
13. Lhuillier, C., Rudqvist, N.-P., Elemento, O., Formenti, S. C. & Demaria, S. Radiation therapy and anti-tumor immunity: exposing immunogenic mutations to the immune system. *Genome Med.* **11**, 40 (2019).
14. Lee, Y. *et al.* Therapeutic effects of ablative radiation on local tumor require CD8+ T cells: changing strategies for cancer treatment. *Blood* **114**, 589–595 (2009).
15. Ngwa, W. *et al.* Using immunotherapy to boost the abscopal effect. *Nat. Rev. Cancer* **18**, 313–322 (2018).
16. Grassberger, C., Ellsworth, S. G., Wilks, M. Q., Keane, F. K. & Loeffler, J. S. Assessing the interactions between radiotherapy and antitumour immunity. *Nat. Rev. Clin. Oncol.* **16**, 729–745 (2019).
17. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).
18. Jin, Y. *et al.* The Diversity of Gut Microbiome is Associated With Favorable Responses to Anti-Programmed Death 1 Immunotherapy in Chinese Patients With NSCLC. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **14**, 1378–1389 (2019).
19. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).

20. Meir, H. van *et al.* Impact of (chemo)radiotherapy on immune cell composition and function in cervical cancer patients. *OncoImmunology* **6**, e1267095 (2017).
21. Jordanova, E. S. *et al.* Human leukocyte antigen class I, MHC class I chain-related molecule A, and CD8+/regulatory T-cell ratio: which variable determines survival of cervical cancer patients? *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **14**, 2028–2035 (2008).
22. Muls, A. *et al.* Systematic Review: The Impact of Cancer Treatment on the Gut and Vaginal Microbiome in Women With a Gynecological Malignancy. *Int. J. Gynecol. Cancer* **27**, 1550–1559 (2017).
23. Gerassy-Vainberg, S. *et al.* Radiation induces proinflammatory dysbiosis: transmission of inflammatory susceptibility by host cytokine induction. *Gut* **67**, 97–107 (2018).
24. Sokol, H. & Adolph, T. E. The microbiota: an underestimated actor in radiation-induced lesions? *Gut* **67**, 1–2 (2018).
25. Lavie, C. J., Osman, A. F., Milani, R. V. & Mehra, M. R. Body composition and prognosis in chronic systolic heart failure: the obesity paradox. *Am. J. Cardiol.* **91**, 891–894 (2003).
26. Lavie, C. J., De Schutter, A. & Milani, R. V. Healthy obese versus unhealthy lean: the obesity paradox. *Nat. Rev. Endocrinol.* **11**, 55–62 (2015).
27. Walker, J. L. *et al.* Laparoscopy Compared With Laparotomy for Comprehensive Surgical Staging of Uterine Cancer: Gynecologic Oncology Group Study LAP2. *J. Clin. Oncol.* **27**, 5331–5336 (2009).
28. S, A.-A., A, S. & J, K. Obesity and cancer. *J. Med.* **33**, 73–86 (2002).
29. Gunderson, C. C., Java, J., Moore, K. N. & Walker, J. L. The impact of obesity on surgical staging, complications, and survival with uterine cancer: A Gynecologic Oncology Group LAP2 ancillary data study. *Gynecol. Oncol.* **133**, 23–27 (2014).
30. Poorolajal, J. & Jenabi, E. The association between BMI and cervical cancer risk: a meta-analysis. *Eur. J. Cancer Prev. Off. J. Eur. Cancer Prev. Organ. ECP* **25**, 232–238 (2016).
31. Nagaraju, G. P. *et al.* The role of adiponectin in obesity-associated female-specific carcinogenesis. *Cytokine Growth Factor Rev.* **31**, 37–48 (2016).
32. Martín-López, R. *et al.* Breast and cervical cancer screening in Spain and predictors of adherence. *Eur. J. Cancer Prev.* **19**, 239–245 (2010).
33. Lacey, J. V. *et al.* Obesity as a potential risk factor for adenocarcinomas and squamous cell carcinomas of the uterine cervix. *Cancer* **98**, 814–821 (2003).
34. Brinton, L. A. *et al.* Risk Factors for Cervical Cancer by Histology. *Gynecol. Oncol.* **51**, 301–306 (1993).
35. Törnberg, S. A. & Carstensen, J. M. Relationship between Quetelet's index and cancer of breast and female genital tract in 47,000 women followed for 25 years. *Br. J. Cancer* **69**, 358–361 (1994).
36. Lane, G. Obesity and gynaecological cancer. *Menopause Int.* **14**, 33–37 (2008).
37. Chen, H.-N. *et al.* The Impact of Body Mass Index on the Surgical Outcomes of Patients With Gastric Cancer: A 10-Year, Single-Institution Cohort Study. *Medicine (Baltimore)* **94**, e1769 (2015).
38. Kohl, K. D., Weiss, R. B., Cox, J., Dale, C. & Denise Dearing, M. Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecol. Lett.* **17**, 1238–1246 (2014).
39. Coryell, M., McAlpine, M., Pinkham, N. V., McDermott, T. R. & Walk, S. T. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. *Nat. Commun.* **9**, 5424 (2018).

40. Millan, B. *et al.* Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent *Clostridium difficile* Infection. *Clin. Infect. Dis.* **62**, 1479–1486 (2016).
41. Wang, Y. *et al.* Fecal microbiota transplantation for refractory immune checkpoint inhibitor-associated colitis. *Nat. Med.* **24**, 1804–1808 (2018).
42. Alexander, J. L. *et al.* Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 356–365 (2017).
43. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
44. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
45. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

507

Table I. Univariate and multivariate Cox regression analysis for recurrence-free survival

Characteristics	Univariate model		Multivariate model	
	HR (95% CI)	***P* value**	**HR (95% CI)**	***P* value**
Age	0.93 (0.87-0.98)	0.0096	0.93* (0.88-0.99)	0.03‡
BMI (kg/m2)	0.92 (0.84-1)	0.096	0.87* (0.77-0.98)	0.02‡
Normal (18.5 to 24.9)	1 (reference)			
Overweight (25 to 29.9)	0.81(0.26-2.53)	0.715	—	—
Obese (30 or more)	0.47(0.13-1.67)	0.240	—	—
Race/Ethnicity				
Asian	1 (reference)		—	—
Black	0.37(0.02-5.90)	0.479	—	—
Hispanic	0.39 (0.05-3.21)	0.382	—	—
White	0.39 (0.05-3.31)	0.390	—	—
Other	4.1309E-08(-inf - +inf)	0.998	—	—
Stage				
I	1 (reference)			
II	1.50 (0.31-7.34)	0.615		
III	3.99 (0.80-20.01)	0.091	—	—
IV	2.54 (0.23-28.12)	0.447	—	—
Grade			—	—
Well	1 (reference)			
Moderate	55297546(-inf - +inf)	0.998		
Poor	97336741.9(-inf - +inf)	0.998		
Unknown	76285161.4(-inf - +inf)	0.998	—	—
Histology				
Squamous	1 (reference)		—	—
Adenocarcinoma/Adenosquamous	1.06(0.34-3.34)	0.918	—	—
Node Level on PET				
Common Iliac	1 (reference)		—	—
External Iliac	1.33 (0.35-4.95)	0.675	—	—
Internal Iliac	0.67 (0.07-6.89)	0.736	—	—

Characteristics	Univariate model		Multivariate model	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Para-Aortic	1.31 (0.14-12.55)	0.818	—	—
None	0.34 (0.06-2.09)	0.245	—	—
Max Tumor Dimension on MRI	1.3 (1-1.8)	0.042	—	—
Smoking status				
Current	1 (reference)		—	—
Former	0.91 (0.10-7.84)	0.934	—	—
Never	0.89(0.11-7.17)	0.909	—	—
Antibiotic Use				
No	1 (reference)			
Yes	78371200.7 (-inf - +inf)	0.998	—	—
Brachytherapy				
HDR	1 (reference)		—	—
PDR	1.41 (0.48-4.149)	0.532	—	—
Baseline Gut Alpha Diversity				
Observed OTU	0.99 (0.97-1)	0.21	—	—
Shannon	0.51 (0.23-1.1)	0.087	0.36* (0.15-0.84)	0.018‡
Simpson	0.025 (0.000036-1.7)	0.087	—	—
Inverse Simpson	0.93 (0.84-1)	0.11	—	—
Fisher	0.95 (0.88-1)	0.23	—	—
Camargo	13 (0.14-1300)	0.27		
Pielou	0.02 (0.00026-1.6)	0.081		

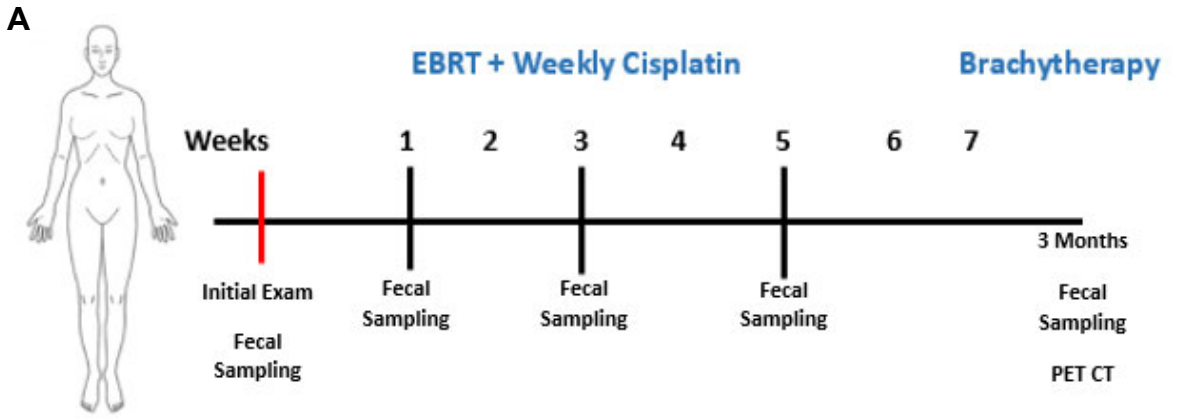
CI, Confidence interval; *HR*, hazard ratio.
∗Significant hazard ratios.
‡Significant *P* value.

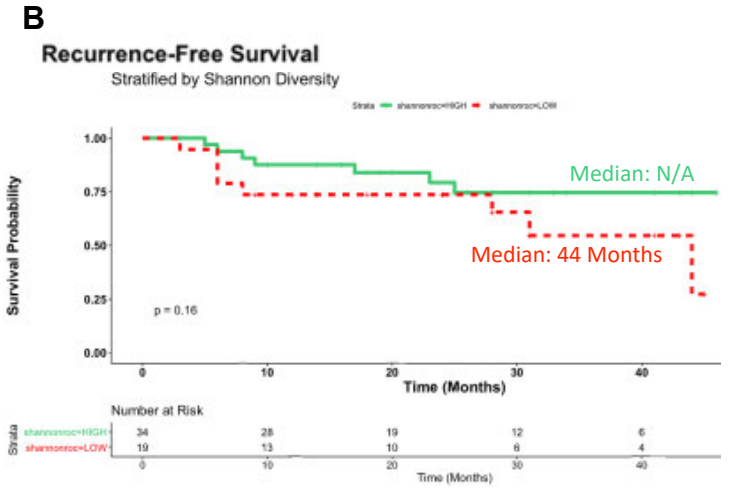
Table II. Univariate and multivariate Cox regression analysis for overall survival

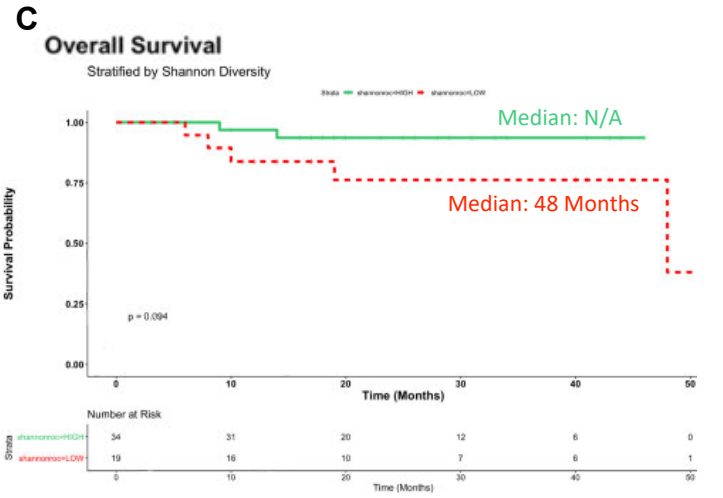
Characteristics	Univariate model		Multivariate model	
	HR (95% CI)	***P* value**	**HR (95% CI)**	***P* value**
Age	0.95 (0.87-1)	0.23	—	—
BMI (kg/m2)	0.83 (0.69-1)	0.055	0.78* (0.623-0.97)	0.025‡
Normal (18.5 to 24.9)	1 (reference)		—	—
Overweight (25 to 29.9)	0.23(0.08-2.32)	0.323	—	—
Obese (30 or more)	0.42 (0.06-4.56)	0.19	—	—
Race/Ethnicity				
Asian	1 (reference)		—	—
Black	4.46E-09 (-inf - +inf)	0.999	—	—
Hispanic	0.23(0.02-2.22)	0.204	—	—
White	0.17 (0.02-1.90)	0.151	—	—
Other	4.48E-09 (-inf - +inf)	0.999	—	—
Stage				
I	1 (reference)			
II	1.19 (0.12-11.43)	0.881		
III	1.49 (0.09-23.93)	0.776	—	—
IV	5.13 (0.32-82.34)	0.248	—	—
Grade			—	—
Well	1 (reference)		—	—
Moderate	116103697.1 (-inf - +inf)	0.999	—	—
Poor	46065187.92(-inf - +inf)	0.999	—	—
Unknown	149251105.9(-inf - +inf)	0.999	—	—
Histology				
Squamous	1 (reference)		—	—
Adenocarcinoma/Adenosquamous	3.40 (0.69-16.90)	0.134	—	—
Node Level on PET				
Common Iliac	1 (reference)		—	—
External Iliac	1.099 (0.09-12.86)	0.306	—	—
Internal Iliac	4.83 (0.24-98.040)	0.999	—	—
Para-Aortic	5.9333E-08 (-inf - +inf)	0.354	—	—

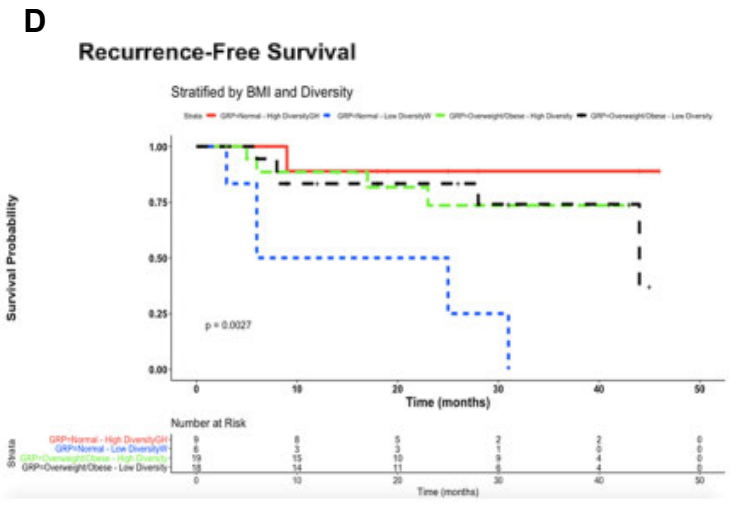
Characteristics	Univariate model		Multivariate model	
	HR (95% CI)	***P* value**	**HR (95% CI)**	***P* value**
None	3.34(0.26-42.69)	0.940	—	—
Max Tumor Dimension on MRI	1.2 (0.77-1.8)	1.2	—	—
Smoking status				
Current	1 (reference)		—	—
Former	106318829.2 (-inf - +inf)	0.999	—	—
Never	61091037.65 (-inf - +inf)	0.999	—	—
Antibiotic Use				
No	1 (reference)			
Yes	0.53 (0.06-4.56)	0.564	—	—
Brachytherapy				
HDR	1 (reference)		—	—
PDR	0.89 (-1.61-1.39)	0.884	—	—
Baseline Gut Alpha Diversity				
Observed OTU	0.98 (0.95-1)	0.14	—	—
Shannon	0.34 (0.1-1.1)	0.08	0.19* (0.043-0.83)	0.028‡
Simpson	0.0059 (1.2e-05-2.9)	0.1	—	—
Inverse Simpson	0.85 (0.7-1)	0.13	—	—
Fisher	0.91 (0.79-1)	0.15	—	—
Camargo	2200 (0.84-5800000)	0.055		
Pielou	0.0036 (5e-06-2.5)	0.093		

CI, Confidence interval; *HR*, hazard ratio.
∗Significant hazard ratios.
‡Significant *P* value.









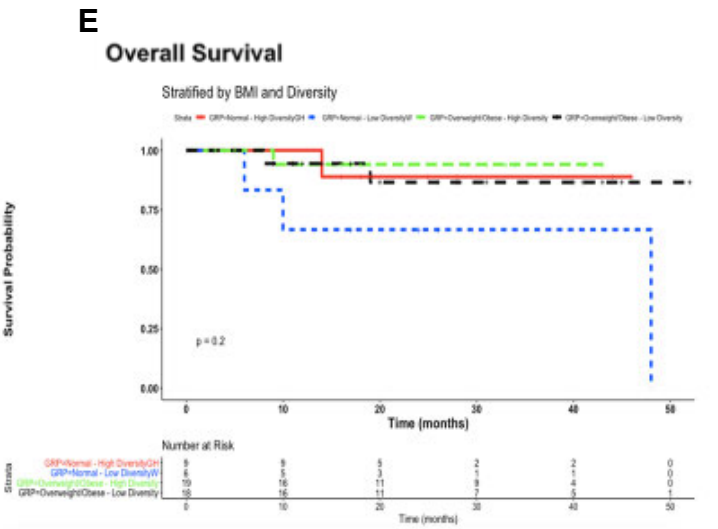


Fig 1. Relationship between gut diversity and BMI. (A) Schema of the sample collection, treatment, and analyses used in the present study. Kaplan-Meier curves for (B) recurrence free survival, (C) overall survival stratified by high and low gut diversity. Kaplan-Meier curves for (D) recurrence free survival, (E) overall survival stratified by BMI and gut diversity. Cases represent patients.

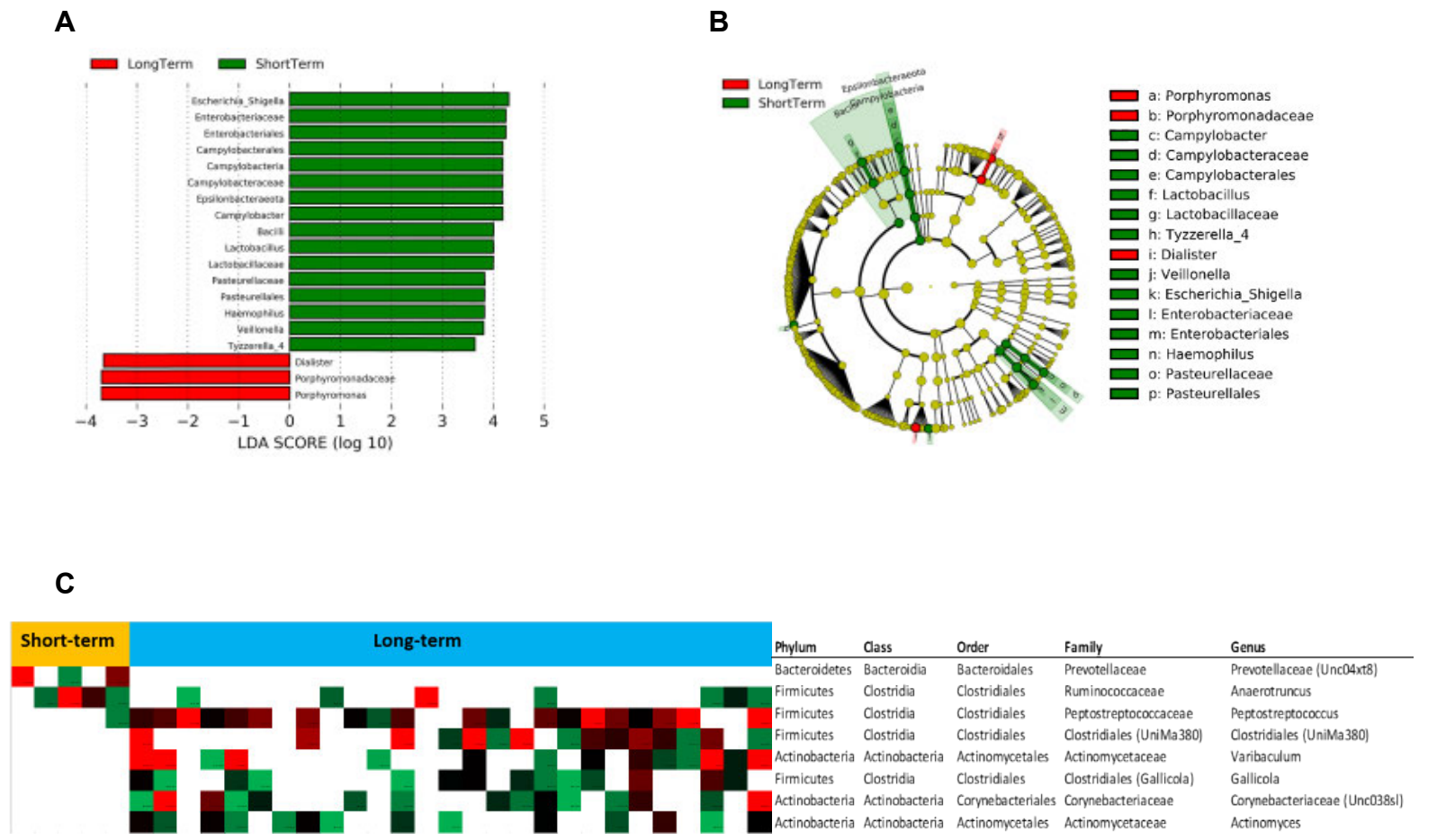


Fig. 2 (A) The different abundance of bacterial genus between the two groups were identified by LEfSe. It was significantly different when alpha value of the factorial Kruskal–Wallis test was <0.05 and the logarithmic LDA score was >3.0. The left histogram showed the LDA scores of genera differentially abundant between the two groups. The taxonomy was listed, followed by its core group. Putative species (Specific OTUs) identified as significantly more enriched/depleted (Fisher/Wilcoxon test p value < 0.05) in patients with short-term vs long-term in baseline samples. (B) Cladogram representation of the significantly different taxa features from phylum (inner circle) to genus (outer circle) (C) The right heatmap showed the relative abundance of specific bacteria by phylum, class, order, family and genus between short-term and long-term survivors.

	P-value correlation of immune metric with baseline gut diversity	R^2	Q value
CD4+Ki.67+ at T4	0.004		0.0714
CD4+CD69+ at T3	0.004		0.1429
CD4+PD1+ at T3	0.0367		0.2143
CD4+CTLA4+ at T3	0.057		0.2857
CD4+		0.1	

Table 3. Correlation of baseline gut diversity (Inverse Shannon Diversity) with phenotype of tumor infiltrating lymphocytes during chemoradiation treatment. The percent of live lymphocytes expressing each markers was correlated with baseline Shannon diversity of the gut microbiome.

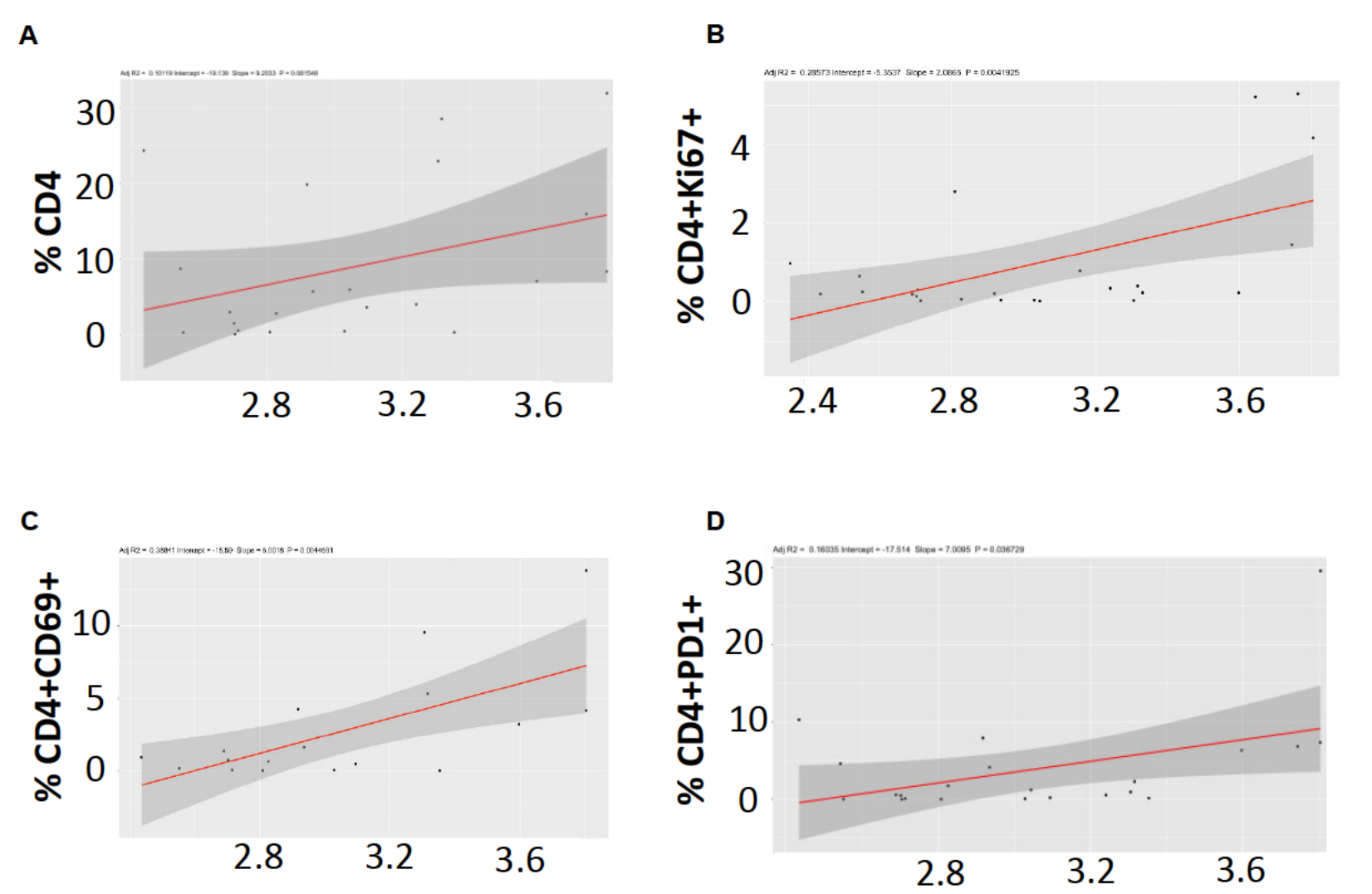
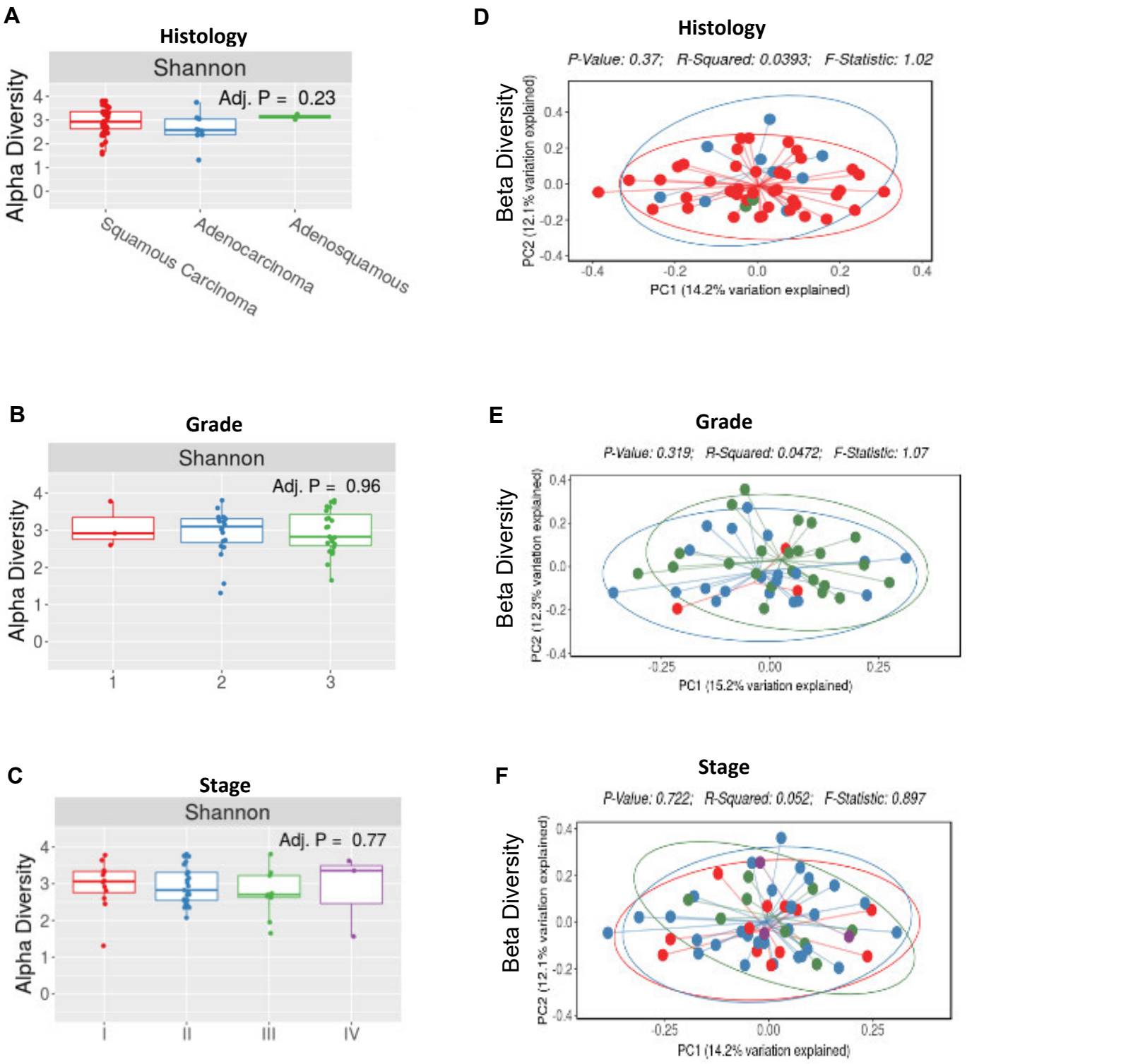
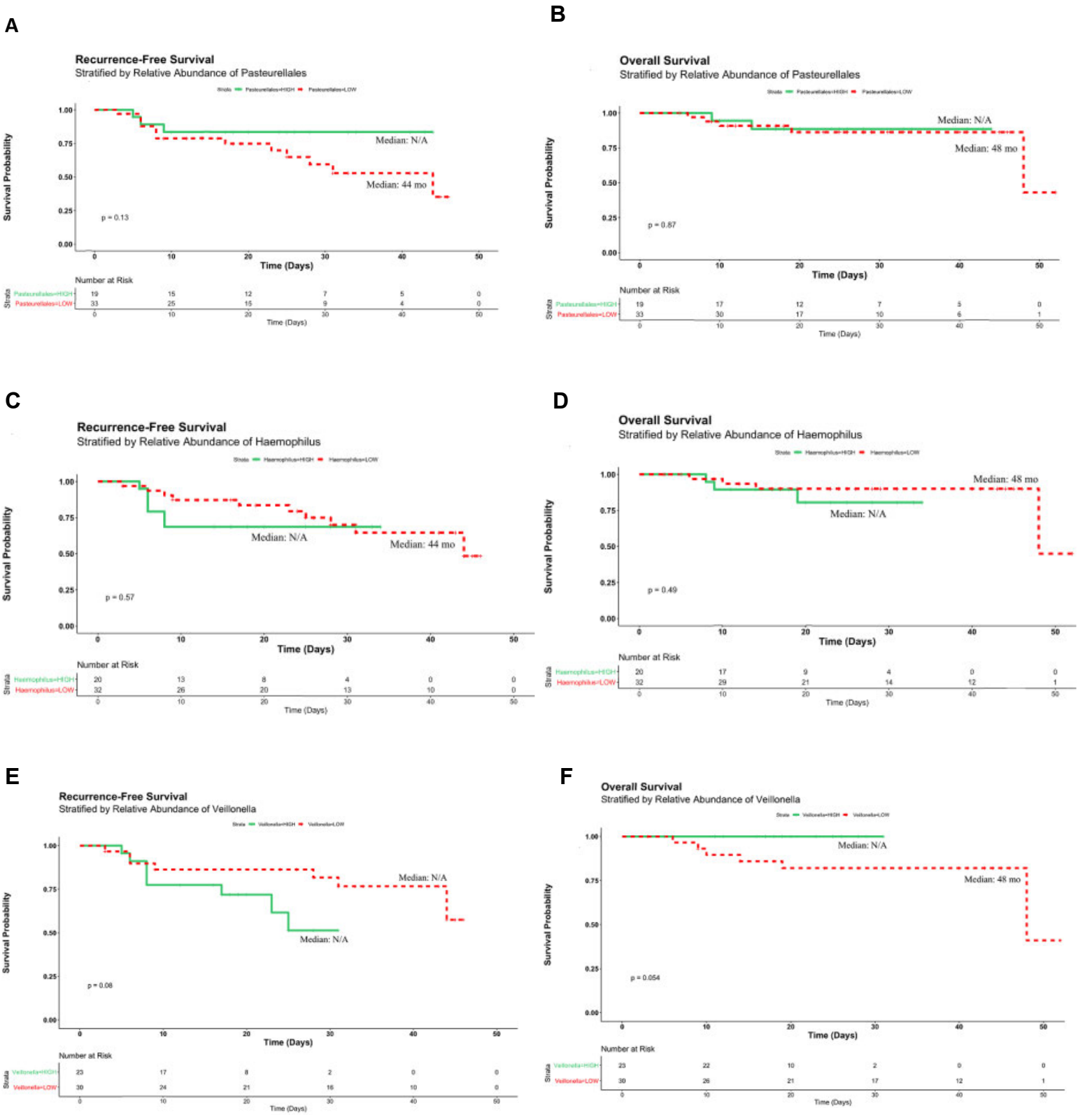


Fig 4. Correlation analysis of Shannon Diversity Index with tumor immune signatures. (A,B,C,D) Spearman correlations between Shannon Diversity Index and CD4, CD4 Ki67, CD4 CD69, and CD4 PD1). Statistical analysis was performed by Spearman correlation or Mann-Whitney tests.

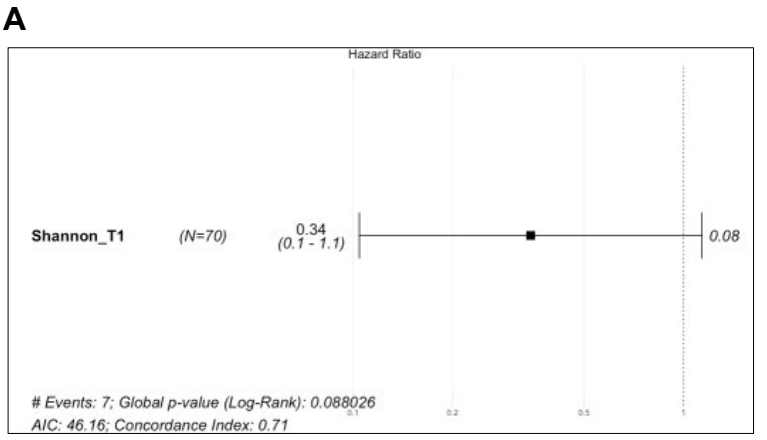
Supplemental Figures

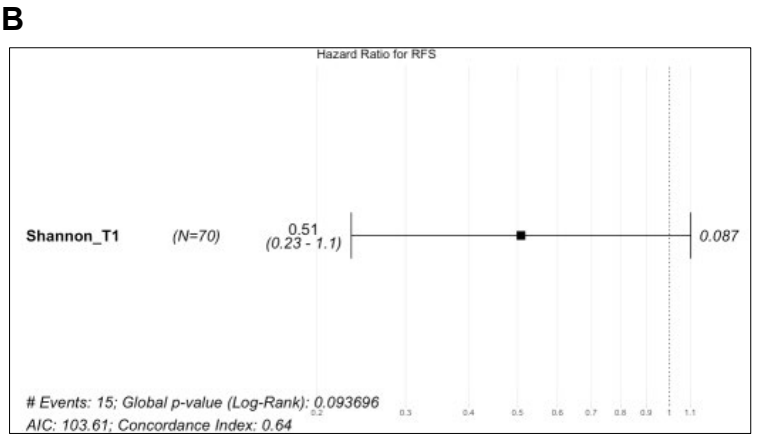


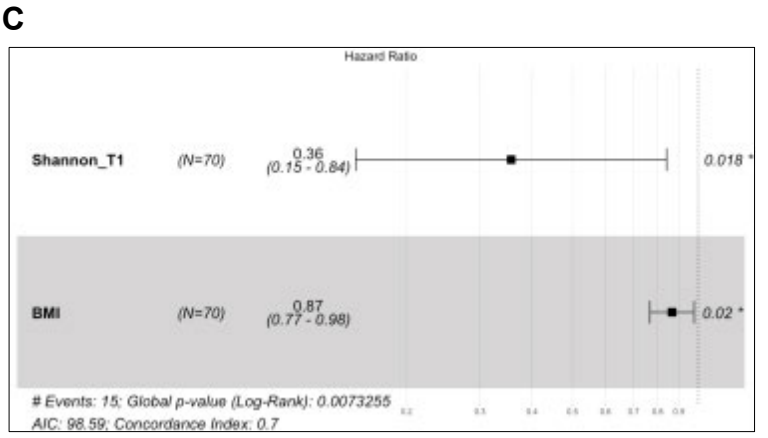
Supplemental Fig 1. The fecal microbiota of individuals with cervical cancer.
The fecal microbiota of individuals with cervical cancer by demographics. Diversity (within sample diversity) was measured using the Shannon diversity metric and Beta diversity (between sample diversity) was determined by unweighted Unifrac. No differences were observed in either metric by cancer histology (A,D), grade (B,E) or cancer stage (C,F).

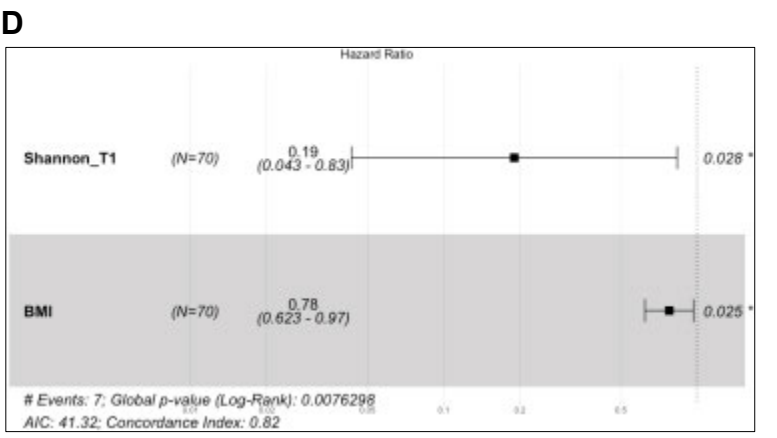


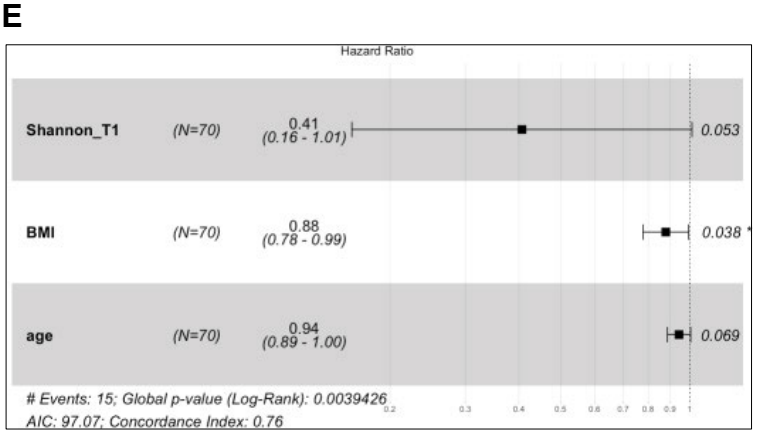
Supplemental Fig 2. Relationship between gut diversity and BMI. Kaplan-Meier curves for (A) recurrence free survival, (B) overall survival stratified by relative abundance of *Pasteurellales*. Kaplan-Meier curves for (C) recurrence free survival, (D) overall survival stratified by relative abundance of *Haemophilus*. Kaplan-Meier curves for (E) recurrence free survival, (F) overall survival stratified by relative abundance of *Veillonella*. Cases represent patients.

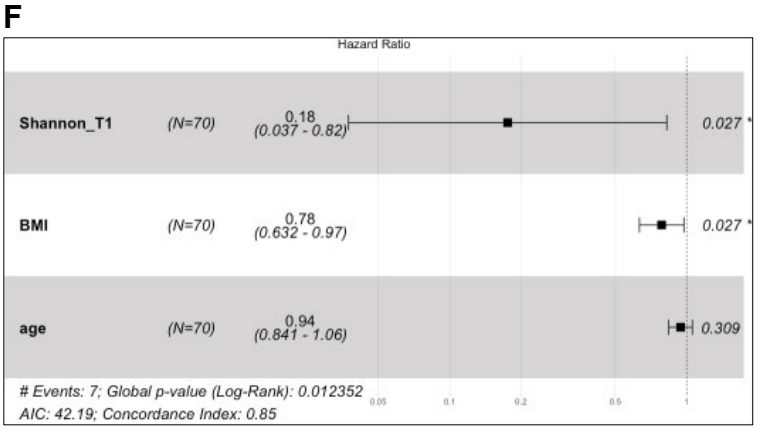












Supplemental Figure 2: Hazard ratio of recurrence according to SDI, BMI and Age

	Observed OTUs	P-value	Shannon	P-Value	Simpson	P-value	Inverse Simpson	P-value	Fisher	P-value
	Coefficient		Coefficient		Coefficient		Coefficient		Coefficient	
Age[1]	0.631	0.181	0.017	0.037*	0.251	0.004*	0.002	0.054	0.128	0.177
BMI[1]	1.167	0.14	0.016	0.237	0.101	0.513	0.002	0.389	0.236	0.14
	Mean±SD		Mean±SD		Mean±SD		Mean±SD		Mean±SD	
Race/Ethnicity[2]										
A	108±79.2		3.07±0.98		12.09±8.86		0.89±0.08		18.8±15.83	
B	87.5±45.14		2.67±0.62		9.1±5.63		0.86±0.06		14.42±9.03	
H	110.55±28.34	0.755	2.89±0.59	0.862	11.23±7.6	0.864	0.86±0.09	0.948	18.76±5.72	0.782
O	87±NA		2.54±NA		6.43±NA		0.84±NA		13.96±NA	
W	109.5±35.8		2.96±0.6		12.12±6.13		0.88±0.09		18.66±7.29	
Smoking History[2]										
Current	161.5±24.24		3.6±0.22		21.24±5.05		0.95±0.01		29.51±5.44	
Former	96.1±34.7	**0.001***	2.77±0.61	**0.033***	9.96±5.6	**0.006***	0.86±0.1	0.152	15.99±6.86	**0.0008***
Never	108.45±28.11		2.9±0.56		11.07±6.57		0.87±0.08		18.34±5.67	
FIGO Stage[2]										
IA1	136±NA		3.33±NA		16.65±NA		0.94±NA		0.94±NA	
IB1	113.6±30.16		3.01±0.39		10.58±4.71		0.89±0.05		0.89±0.05	
IB2	105.5±32.76		2.86±0.87		12.1±8.84		0.85±0.15		0.85±0.15	
IIA	81±5.29	0.841	2.65±0.09	0.9	7.95±1.55	0.835	0.87±0.03	0.693	0.87±0.03	0.84
IIB	111.12±37.19		2.98±0.53		11.87±6.47		0.89±0.06		0.89±0.06	
IIIB	106.22±34.7		2.74±0.66		9.49±7.2		0.84±0.12		0.84±0.12	
IVA	105.67±52.73		2.85±1.12		14.12±10.69		0.85±0.16		0.85±0.16	
Histology[2]										
Adenocarcinoma	93.67±30.83		2.63±0.66		8.68±6.5		0.83±0.11		15.45±6.04	
Adenosquamous	117.5±10.61	0.372	3.13±0.15	0.285	13.52±2.59	0.398	0.92±0.01	0.219	20.02±2.18	0.377
Squamous carcinoma	111.05±35.67		2.96±0.58		11.84±6.76		0.88±0.08		18.97±7.24	
Grade[2]										
1	110±22.87		3.1±0.61		14.23±11.48		0.89±0.07		18.56±4.67	
2	108.95±37.5	0.78	2.92±0.63	0.537	11.41±5.61	0.606	0.88±0.1	0.729	18.57±7.55	0.763
3	111.43±36.41		2.96±0.58		11.8±6.94		0.88±0.09		19.06±7.35	
Unknown	95.71±25.88		2.61±0.56		8.56±7.08		0.84±0.07		15.78±5.25	
Node level on PET[2]										
Common Iliac	121±35.34		3.01±0.76		13.13±7.51		0.87±0.13		0.87±0.13	
External Iliac	113.33±34.87		3±0.47		12.28±6.61		0.89±0.05		0.89±0.05	
Internal Iliac	121.2±44.89	0.188	3.06±0.86	0.497	11.79±5.55	0.4526	0.87±0.12	0.611	0.87±0.12	0.18
Para-Aortic	98±34.7		2.93±0.74		12.74±12.56		0.86±0.09		0.86±0.09	
None	91.53±26.95		2.67±0.53		8.61±5.18		0.85±0.09		0.85±0.09	
Brachytherapy[3]										
HDR	101.6±32.31	0.274	2.77±0.55	0.191	9.11±4.63	**0.033***	0.86±0.09	0.243	17.03±6.52	0.266
PDR	112.42±35.79		2.99±0.61		12.73±7.36		0.88±0.09		19.25±7.23	
Antibiotic use during treatment[3]										
No	113±29.17	0.755	2.86±0.57	0.866	8.76±6.02	0.362	0.85±0.08	0.448	19.23±6.13	0.786
Yes	107.85±35.36		2.91±0.6		11.64±6.72		0.88±0.09		18.33±7.13	

[1]Simple linear regression; [2]ANOVA; [3]Independant t-test

Supplemental Table 1: Patient and tumor characteristics (N=55)		
	n	(%)
Median age, yrs (range)	48 (28-72)	—
BMI, Mean (SD), kg/m2	28.7(6.06)	—
Race/Ethnicity		
Asian	2	(36.4)
Black	4	(18.2)
Hispanic	24	(43.6)
White	24	(43.6)
Other	1	(1.8)
FIGO Stage		
IA1	1	(1.8)
IA2	0	(0)
IB1	5	(9.09)
IB2	6	(10.9)
IIA	3	(5.45)
IIB	28	(50.9)
IIIA	9	(16.3)
IIIB	0	(0)
IVA	3	(5.45)
IVB	0	(0)
Grade		
Well	4	(7.2)
Moderate	20	(36.3)
Poor	25	(45.4)
Unknown	6	(10.9)
Histology		
Squamous	43	(78.1)
Adenocarcinoma	8	(18.1)
Adenosquamous	3	(3.63)
Node Level on PET		
Common Iliac	9	(16.3)

External Iliac	23	(41.8)
Internal Iliac	5	(9.09)
Para-Aortic	3	(5.45)
None	15	(27.2)
Median cervical tumor size (cm)	5.4	—
Smoking status		
Current	4	(7.27)
Former	20	(36.3)
Never	31	(56.3)
Antibiotic Use		
No	5	(9.1)
Yes	50	(90.9)
Brachytherapy		
HDR	21	(38.2)
PDR	34	(61.8)
Concurrent Chemotherapy (cycles)		
Cisplatin		
(1-3)	2	(3.6)
(≥4)	51	(92.7)
Carboplatin		
(2)	1	(1.8)
Carboplatin + Cisplatin		
(2)+(2)	1	(1.8)

FIGO- International Federation of Gynecology and Obstetrics
HDR-High Dose Rate
PDR- Pulsed Dose Rate

Supplemental Table II. Univariate Cox regression analysis for recurrence-free survival – Alpha Diversity all time points

Characteristics	Univariate model	
	HR (95% CI)	***P* value**
Observed OUT		
Time Point 1	0.99 (0.97-1)	0.21
Time Point 2	1 (0.97-1)	0.69
Time Point 3	0.99 (0.96-1)	0.59
Time Point 4	1 (0.98-1)	0.71
Time Point 5	1 (0.98-1)	0.77
Shannon		
Time Point 1	0.51 (0.23-1.1)	0.087
Time Point 2	0.94 (0.2-4.4)	0.94
Time Point 3	1.2 (0.25-5.6)	0.83
Time Point 4	0.83 (0.35-1.9)	0.66
Time Point 5	2.7 (0.13-57)	0.51
Simpson		
Time Point 1	0.025 (0.00036-1.7)	0.087
Time Point 2	13 (1.4e-05-1.2e+07)	0.13
Time Point 3	52 (6.6e-05-4.1e+07)	0.57
Time Point 4	0.31 (0.013-7.8)	0.48
Time Point 5	130000 (7.5e-13-2.2e+22)	0.56
Inverse Simpson		
Time Point 1	0.93 (0.84-1)	0.11
Time Point 2	0.96 (0.79-1.2)	0.69
Time Point 3	1 (0.95-1.1)	0.34
Time Point 4	1 (0.92-1.2)	0.54
Time Point 5	1.1 (0.81-1.4)	0.59
Fisher		
Time Point 1	0.95 (0.88-1)	0.23
Time Point 2	0.97 (0.86-1.1)	0.66
Time Point 3	0.96 (0.83-1.1)	0.6

Characteristics	Univariate model	
	HR (95% CI)	***P* value**
Time Point 4	1 (0.91-1.2)	0.69
Time Point 5	1 (0.89-1.2)	0.81

CI, Confidence interval; *HR*, hazard ratio.
∗Significant hazard ratios.
‡Significant *P* value.

Supplemental Table III. Univariate Cox regression analysis for overall survival – Alpha Diversity all time points

Characteristics	Univariate model	
	HR (95% CI)	*P* value
Observed OUT		
Time Point 1	0.98 (0.95-1)	0.14
Time Point 2	0.98 (0.94-1)	0.35
Time Point 3	0.97 (0.92-1)	0.21
Time Point 4	1 (0.96-1)	0.98
Time Point 5	NA (NA-NA)	1
Shannon		
Time Point 1	0.34 (0.1-1.1)	0.08
Time Point 2	0.48 (0.063-3.7)	0.48
Time Point 3	1.2 (0.25-5.6)	0.83
Time Point 4	0.23 (0.037-1.4)	0.11
Time Point 5	NA (NA-NA)	1
Simpson		
Time Point 1	0.0059 (1.2e-05-2.9)	0.1
Time Point 2	0.45 (1.1e-08-1.8e+07)	0.93
Time Point 3	0.009 (1.7e-07-490)	0.4
Time Point 4	1.4 (0.00063-3200)	0.93
Time Point 5	NA (NA-NA)	1
Inverse Simpson		
Time Point 1	0.85 (0.7-1)	0.13
Time Point 2	0.86 (0.62-1.2)	0.39
Time Point 3	0.81 (0.61-1.1)	0.15
Time Point 4	0.89 (0.66-1.2)	0.46
Time Point 5	NA (NA-NA)	1
Fisher		
Time Point 1	0.91 (0.79-1)	0.15
Time Point 2	0.91 (0.74-1.1)	0.34
Time Point 3	0.84 (0.64-1.1)	0.22

Characteristics	Univariate model	
	HR (95% CI)	***P* value**
Time Point 4	0.99 (0.79-1.3)	0.94
Time Point 5	NA (NA-NA)	1

CI, Confidence interval; *HR*, hazard ratio.
∗Significant hazard ratios.
‡Significant *P* value.

Date : 4/14/2020 1:12:13 PM
From : "Sims,Travis T." TTSims@mdanderson.org
To : "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Ning,Matthew Stephen" MSNing@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org, "Court,Kyoko" KCourt1@mdanderson.org, "Wu,Xiaogang" XWu10@mdanderson.org, "Mezzari, Melissa" [redacted] "Delgado Medrano,Andrea Yizel" AYDelgado@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Solley,Travis N" TNSolley@mdanderson.org, "Ahmed-Kaddar,Mustapha" MAhmed10@mdanderson.org, [redacted] "Schmeler,Kathleen M" KSchmele@mdanderson.org, [redacted]
Cc : "Klopp,Ann H" AKlopp@mdanderson.org, "Biegert, Greyson" [redacted]
Subject : Manuscript - Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients
Attachment : Manuscript - Cervical Tumor Microbiome Botswana 4-14-20.docx;Botswana Manuscript - Table 1..docx;Botswana Manuscript - Table 2..docx;Botswana Manuscript Figure 1.pptx;Botswana Manuscript Figure 2.pptx;

Hello all,

We have completed the first draft of the manuscript for our project "Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients".

You have been included on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to The *International Journal of Gynecological Cancer* (*IJGC*).

Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.

I am grateful for your time and feedback regarding this project! We hope to submit by 4/24/20.

Best,
Travis

Travis T. Sims, MD, MPH
Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

Title: Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients

Authors and Affiliations

Travis T. Sims, MD, MPH[1*], Greyson W.G. Biegert, BS[2*], Kebatshabile Ngoni, BSc[3], Molly B. El Alam, MPH[2], Matthew S. Ning, MD, MPH[2], Tatiana V. Karpinets, PhD[4], Kyoko Yoshida-Court, PhD[2], Xiaogang Wu, PhD[4], Andrea Y. Delgado Medrano, BS[2], Melissa P. Mezzari, PhD[5], Joseph Petrosino, PhD[5], Nadim J. Ajami, PhD[4], Travis Solley, BS[2], Mustljapha Ahmed-Kaddar, BS[2], Kathleen M. Schmeler, MD[1], Lauren E. Colbert, MD, MSCR[2], Stephen Hahn, MD[2], Ann H. Klopp, MD, PhD[2], Surbhi Grover, MD, MPH[6]

[1]Department of Gynecologic Oncology and Reproductive Medicine, Division of Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX

[2]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

[3]Department of Biological Sciences, University of Botswana, Gaborone, Botswana

[4]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX

[5]Department of Molecular Virology and Microbiology, Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX

[6]Department of Radiation Oncology, Botswana-University of Pennsylvania Partnership, Gaborone, Botswana, and Philadelphia, PA

*These authors contributed equally.

Corresponding Author

Travis T. Sims, MD, MPH

Department of Gynecologic Oncology and Reproductive Medicine, Division of Surgery

The University of Texas MD Anderson Cancer Center

1155 Herman Pressler, CPB 6.3258, Unit 1362

Houston, TX 77030

Phone: 713-563-4578

Fax: 713-745-2398

Email: ttsims@mdanderson.org

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Research Support

This research was supported in part by the National Institutes of Health (NIH) through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant 5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK).

Role of Funding Sources

47 The funding sources were not involved in the research hypothesis development, study design,
48 data analysis, or manuscript writing. Data access was limited to the authors of this manuscript.

ABSTRACT

Introduction We characterized the cervical 16S rDNA microbiome of cervical dysplasia and locally advanced cervical cancer in patients in Botswana. **Methods** Our prospective study included 31 patients (21 with dysplasia and 10 with cancer). We used the Shannon diversity index to evaluate alpha (within sample) diversity and UniFrac (weighted and unweighted) and Bray-Curtis distances to evaluate beta (between sample) diversity. We compared the relative abundance of microbial taxa between samples using linear discriminant analysis effect size. **Results** Alpha diversity was significantly higher in cervical cancer patients than in cervical dysplasia patients ($p<0.05$). Beta diversity (weighted UniFrac Bray-Curtis, $p<0.01$) also significantly differed. The results of linear discriminant analysis effect size demonstrated that multiple taxa significantly differed between cervical dysplasia and cancer patients. *Lachnospira* bacteria, in the *Clostridia* class, were significantly enriched in cervical dysplasia patients, while *Proteobacteria*, members of the *Firmicutes* phyla and the *Comamonadaceae* family were enriched in cervical cancer patients. **Discussion** The results of our study suggest that differences exist in the diversity and composition of the cervical microbiota between cervical dysplasia and cervical cancer patients in Botswana. Additional studies are needed to validate these findings in larger cohorts to determine the biological significance of these observed differences in women living in southern Africa.

Keywords: Cervical dysplasia; cervical cancer; gynecologic cancer; cervical microbiota; microbiome; HIV; Botswana

Highlights:

- In this cohort of women in Botswana, cervical microbiome diversity was higher in cervical cancer patients than in cervical dysplasia patients.
- The cervical microbiota of women with cervical cancer have a distinct composition compared with those of women with cervical dysplasia.
- Currently, there is an important gap in the number of studies investigating the cervical microbiome and gynecologic cancers in sub-Saharan African patients.

INTRODUCTION

Cervical cancer is one of the most common malignancies globally and the most common cause of cancer death among African women[1]. More than half a million new cases of invasive cervical cancer are expected to be diagnosed worldwide in 2020, resulting in more than 300,000 deaths[2]. African women have a far higher risk of cervical cancer than do women in regions with more access to preventative health care screening[1]. Fourteen percent of the world's cervical cancer cases and 18% of cervical cancer-related deaths occur in women living in sub-Saharan Africa[1,3]. The incidence of cervical cancer in southern Africa, which includes the countries of Botswana, Lesotho, Namibia, South Africa, and Swaziland, is expected to increase by roughly 35% by 2030[1].

It is well established that persistent exposure to the human papilloma virus (HPV) is an antecedent to cervical cancer[4]. Women with HIV are at increased risk of HPV infection and ultimately, cervical cancer, despite access to anti-retroviral therapy[5]. The high regional prevalence of HIV in countries such as Botswana underscores the importance of cervical cancer prevention in these regions. Botswana established one of the original nationwide HIV treatment programs[6] in Africa, but despite a corresponding decline in HIV-associated mortality, the incidence of cervical cancer remains among the highest globally (36.6 per 100,000), with nearly two-thirds of cases occurring in HIV-positive women[7].

The microbiome has recently been demonstrated to play a critical role in cancer progression and metastasis and cancer-directed therapy response[8]. The female cervix is a microbiome-rich environment, but the effect of this microbiome on cervical cancer development and progression is limited and not well understood[9]. Given the expected incidence of cervical

cancer in 2020, understanding the effect of the cervical flora on cancer progression and response, as well as the converse effect of treatments such as chemoradiation therapy, represents a critical unmet need, especially in vulnerable populations, such as women residing in Botswana.

To our knowledge, no published studies exist that specifically explore the cervical tumor microbiome in women in Botswana. Cervical cancer is uniquely positioned for such a crucial investigation, as it allows direct visualization and contact with the primary tumor at the initiation of treatment.

Because cervical microbial differences can affect cervical cancer risk and treatment through several pathways, we characterized the 16S rDNA cervical microbiome of women with cervical dysplasia and locally advanced cervical cancer in Botswana. We hypothesize that the cervical microbiome of cervical cancer patients is distinct from that of dysplasia patients. We theorize that the longitudinal identification of persistent bacterial strains that are associated with the cervical microbiome will allow us to further study the organisms that stably colonize cervical cancers, detect bacterial strains that are associated with treatment response, and lay the groundwork for developing interventions that alter the tumor microbiota to improve cancer outcomes.

PATIENTS AND METHODS

Participants and Clinical Data

We prospectively identified patients with newly diagnosed, biopsy-proven cervical dysplasia or locally advanced, non-metastatic cervical carcinoma who were treated at the University of Botswana General Hospital oncology clinic between July 24, 2018, and February 22, 2019. The

study protocol, the final approved informed consent document and the subject recruitment information were submitted to the Institutional Review Board (IRB) and samples used for this study were obtained following ethical approval by the IRB at the University of Botswana [IRB reference number: UBR/RES/IRB/BIO/045], the University of Pennsylvania (IRB reference number: 830039), and the University of Texas MD Anderson Cancer Center (IRB reference number: MDACC 2014-0543). The subject's informed consent was mandatory for study participation and was obtained in writing.

Patient ineligibility criteria included incident or prevalent cancer other than cervical cancer and currently pregnant women. Medical history and current medication use were assessed via an in-person interview with a clinical provider or trained study staff. We reviewed patients' medical records to obtain demographic and clinico-pathologic data. All cancer patients had a new diagnosis of locally advanced, non-metastatic carcinoma of the cervix and underwent definitive chemoradiation (CRT) with external beam radiation therapy followed by brachytherapy, but samples used for this study were collected prior to any cancer therapy.

Sample Collection and DNA Extraction

Cervical samples were collected using a matrix-designed quick-release Isohelix swab. The swabs were placed in 20 μL of protease K and 400 μL of lysis buffer (Isohelix) and stored at –80°C within 1 hour of sample collection. Bacterial genomic DNA was extracted using a MO BIO PowerSoil DNA Isolation Kit (MO BIO Laboratories). Samples were shipped to the US for downstream applications that include DNA processing and sequencing.

16S rRNA Gene Sequencing and Sequence Data Processing

16S rRNA gene sequencing of the cervical swabs was performed at the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine (Houston, Texas) using methods adapted from those used for the Human Microbiome Project.[10] The 16S rDNA V4 region was amplified by PCR using primers that contained sequencing adapters and single-end barcodes, allowing the pooling and direct sequencing of PCR products. Amplicons were sequenced on the MiSeq platform (Illumina) using the 2x250-bp paired-end protocol, yielding paired-end reads that overlapped almost completely. The sequence reads were de-multiplexed, quality filtered, and subsequently merged using USEARCH version 7.0.1090 (4). 16S rRNA gene sequences were clustered into OTUs at a similarity cut-off value of 97% using the UPARSE algorithm.[11] To generate taxonomies, we mapped OTUs to an optimized version of the SILVA rRNA database containing the 16S v4 region. A custom script was used to construct an OTU table from the output files generated, as described above, for downstream analyses of alpha diversity, beta diversity, and phylogenetic trends. Principal coordinates analysis was performed by institution and sample set to ensure that no batch effects were present.

Statistical Analyses

For the microbiome analysis, the rarefaction depth was set at 3651 reads. Alpha (within sample) diversity was examined using the Shannon diversity index, and beta (between sample) diversity was examined using UniFrac (weighted and unweighted) and Bray-Curtis distances. We

compared the relative abundance of microbial taxa and genera between samples; we then determined differentially abundant bacterial genera by case status using linear discriminant analysis (LDA) effect size (LEfSe),[12] applying the 1-against-all strategy with a threshold of 4 on the logarithmic LDA score for discriminative features and an α of 0.05 for the factorial Kruskal-Wallis test among classes. LEfSe was restricted to bacteria that were present in 20% or more of the study population. Observed differences were subjected to paired analysis using two sample Z test for proportions, or Student t test where appropriate.

RESULTS

We characterized the 16S rDNA cervical microbiome in 31 cervical dysplasia and cancer patients (21 with dysplasia and 10 with cancer). Clinico-pathologic data for all patients are summarized in Table 1. Cervical dysplasia patients were classified according to their histological grade of cervical intraepithelial neoplasia ([CIN] stage I-III). Approximately 58% of the patients in the study (18 of 31) had CIN stage III, and approximately 32% (10 of 31) had cervical cancer (in all cases, squamous cell cancer with moderate or poor differentiation). HPV status was unknown in all patients at the time of cervical sampling.

We first analyzed patients' microbiota with respect to HIV status. Neither α diversity ($p=0.8$) nor β diversity ($p=0.19$) varied by HIV status (Figure 1A,B), and the top 10 most abundant genera were similar among all cervical cancer patients (Figure 1C), suggesting that bacterial taxa dominance does not vary by HIV status.

We then sought to extend our analysis to characterize variations in the cervical microbiome by cervical dysplasia vs cervical cancer. Patients' clinical and demographic

characteristics are displayed in Table 2. The mean age and BMI were similar between cervical dysplasia patients and cervical cancer patients (mean age, 41.8 vs 50.7 years [p=0.1], and mean BMI, 26.3 vs. 30.0 kg/m^2 [p=0.19], respectively). We observed a statistically significant higher α diversity, as measured by SDI ($p<0.05$), in cervical dysplasia patients than in cervical cancer patients (Figure 2A). CIN III patients tended to have higher α diversity than did CIN II patients (Figure 2B). As with α diversity, overall β diversity differed significantly by cancer status (weighted Bray-Curtis Unifrac; $p<0.01$) (Figure 2C,D). The top 10 most abundant genera in cervical samples were similar among all cervical dysplasia and cervical cancer patients (Figure 2E). The percentage of subjects with a cervical microbiome dominated by *Lactobacillus* was low in both groups but lower in the cervical cancer cohort (1 of 10 patients).

We used LEfSe to identify the bacterial genera that were differentially enriched in our cohort of patients ($p<0.05$, LDA score >2). We found that the genera *Ersipelotrichia*, *Erysipelotrichales*, *Erysipelotrichaceae*, and *Ruminiclostridium* were enriched in HIV-positive patients, while only *Filifactor* was significantly enriched in HIV-negative patients (Figure 1D,E). We found that the genus *Lachnospira*, in the *Clostridia* class of bacteria, was significantly enriched in cervical dysplasia patients, while several *Proteobacteria* taxa (*Betaproteobacteria*, *Gammaproteobacteria*, and *Burkholderiaceae*) and members of the *Firmicutes* phyla (*Erysiopelotrichaceae* and *Synergistaceae*) and the *Comamonadaceae* family were significantly enriched in cervical cancer patients ($p<0.05$, LDA score >2) (Figure 2F,G).

DISCUSSION

In this study, we characterized the cervical microbiome of cervical dysplasia and cervical cancer patients living in Botswana. We hypothesized that the microbiome of cervical cancer patients would be distinct from that of dysplasia patients. We observed significant differences in cervical α and β diversity between these groups of patients, as well as compositional differences. The results of an overall analysis of α and β diversity revealed that the groups did not differ in regard to HIV status.

The influence of the cervical cancer microbiome site throughout treatment is poorly understood. Research has focused on exploring the relative abundance of bacteria in the vaginal epithelium, with the assignment of community-state types based on the richness of *Lactobacilli* species[13-15]. The presence and abundance of specific *Lactobacilli* species, for example, *L. crispatus*, *L. gasseri*, or *L. jensenii*, is thought to be associated with a predisposition to bacterial vaginosis (BV) and other pro-inflammatory states[16,17].

However, despite the comparative wealth of data focused on the vaginal microbiome, the ectocervical microbiome has yet to be well described. Most studies have concentrated on characterizing it in the setting of pregnancy or pelvic inflammatory disease. Previous studies using 16S rDaNA sequencing have suggested that in pregnancy, cervical microbiota diversity differs by race[18] and that the presence of non-*Lactobacillus* community state types is associated with a robust cervical inflammatory response in the setting of pre-term, premature membrane rupture[19,20]. Wang et al. demonstrated that in patients with pelvic inflammatory disease, the cervical microbiota is dominated by *Lactobacillus* and *Gardnerella*, again suggesting that the abundance of these different taxa is associated with both acute and chronic inflammatory states[21]. It is thought that these states of polybacterial dysbiosis and chronic local inflammation

encourage the perseverance of HPV, which ultimately promotes the development of cervical dysplasia and carcinogenesis in the setting of persistent HPV exposure[15,17,22-25].

Persistent HPV infections are thought to trigger an innate immune response, resulting in the suppression of infected cervicovaginal mucosal cells[16,26,27]. An altered mucosal microenvironment leads to the growth of anaerobic organisms at the expense of *Lactobacillus* growth, creating cervicovaginal dysbiosis[28]. LEfSe was designed to detect bacterial taxa that are associated with a specific state[29]. In our study, LEfSe identified *Clostridia*, *Firmicutes*, and *Lachnospira* as taxa that were negatively associated with cervical cancer and several *Proteobacteria* as taxa that were positively associated with cervical cancer compared with cervical dysplasia.

Dysbiosis causes cervicovaginal inflammation and other unfavorable changes in the cervicovaginal mucosal barrier. Worldwide, the most common type of cervicovaginal dysbiosis, which is defined as a cervicovaginal microbiome that is not dominated by *Lactobacilli*, is BV[30]. BV is characterized by a persistent decrease in *Lactobacilli* and an increase in fastidious anaerobes[26]. Globally, the prevalence of BV is highest in women living in sub-Saharan Africa and in women of sub-Saharan African descent[30]. Cervicovaginal dysbiotic states, such as BV, lead to an altered metabolic profile and reduced cervicovaginal barrier function. This dysbiotic state is not only associated with an increased acquisition of HIV, but also with high-risk HPV, cervical dysplasia, and ultimately cervical cancer[26,31]. The percentage of subjects with their cervical microbiome dominated by *Lactobacillus* was low in our cohort of patients. The proportion of dysplasia patients with *Lactobacillus*-dominated cervical microbiomes was higher than that of cancer patients. The lack of *Lactobacilli* identified in our cervical dysplasia and cervical cancer patients supports this rationale and suggests that cervicovaginal microbes are

important in preventing or enhancing the acquisition and pathogenesis of HPV and HIV. Identifying the microbes that are associated with enhanced pathogenesis and ultimately oncogenesis or tumorigenesis is especially important in susceptible populations such as HIV-positive women in Botswana. Historically, microbiome cervical cancer research has been limited to mainly Western industrialized populations. We hope that our findings in women in Botswana provide a timely and critical glimpse into this uniquely vulnerable population.

The gut microbiome and its influence on carcinogenesis and prognosis has been well described, most notably in melanoma and colorectal cancer[8,32,33]. Bullman et al. recently identified colonization by *Fusobacterium* and its associated microbiome *Bacteriodes*, *Selenomas*, and *Prevotella* at both the primary tumor and the distant paired metastatic site in colorectal cancer. Thus, it is possible that the colonized organisms that inhabit the primary tumor migrate with primary tumor cells to distant locations and manipulate microbiota diversity at sites, ultimately leading to poor anti-tumor immunity[34]. Identifying the specific organisms that colonize the tumor microbiota will provide further insight into the mechanisms that modulate immune response and potentiate tumor cell growth[31].

Although the present study yielded intriguing findings, it was limited by its small sample size. We acknowledge this possible limitation, but our sample size is suggestive of the complexity associated with using 16S rDNA next-generation sequencing to evaluate the cervical microbiome in a remote population; complete data collection was limited, and field circumstances were challenging. Our study design also prevents us from determining the causal associations or mechanisms that are associated with differences in the cervical microbiota and cervical dysplasia or cancer; this is an area that deserves further study. These limitations are

unlikely to fully explain the large differences that we observed between cervical dysplasia and cancer patients.

In conclusion, our study demonstrated hypothesis-generating differences in the cervical microbial profiles of Botswana cervical cancer patients compared to those of cervical dysplasia patients. The lack of *Lactobacilli* in our samples supports the rationale that cervicovaginal dysbiotic states, which are characterized by a persistent decrease in *Lactobacilli*, are associated with a higher incidence of HIV, cervical dysplasia, and cervical cancer. We anticipate that our findings will help improve our understanding of the essential functional role of the tumor microbiome in cervical cancer. Additional studies are needed to validate these findings in larger cohorts and to determine the biological significance of these observed differences in women living in southern Africa.

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Author Contributions

All authors were involved with subject identification and data collection, interpretation of the statistical analysis, and review and approval of the final manuscript. The study concept was developed by LEC, AK, GWGB, and TTS. GWGB and TTS helped draft the manuscript.

Acknowledgements

This work was supported in part by the National Institutes of Health through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant #5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK). We gratefully acknowledge the patients who participated in this study.

REFERENCES

1. De Vuyst H, Alemany L, Lacey C, et al. The burden of human papillomavirus infections and related diseases in sub-saharan Africa. *Vaccine*. 2013;31 Suppl 5:F32-46. doi:10.1016/j.vaccine.2012.07.092

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34. doi:10.3322/caac.21551

3. Klein C, Kahesa C, Mwaiselage J, West JT, Wood C, Angeletti PC. How the Cervical Microbiota Contributes to Cervical Cancer Risk in Sub-Saharan Africa. *Front Cell Infect Microbiol*. 2020;10. doi:10.3389/fcimb.2020.00023

4. Hammer A, Rositch A, Qeadan F, Gravitt PE, Blaakaer J. Age-specific prevalence of HPV16/18 genotypes in cervical cancer: A systematic review and meta-analysis. *Int J Cancer*. 2016;138(12):2795-2803. doi:10.1002/ijc.29959

5. Dryden-Peterson S, Bvochora-Nsingo M, Suneja G, et al. HIV Infection and Survival Among Women With Cervical Cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2016;34(31):3749-3757. doi:10.1200/JCO.2016.67.9613

6. Gaolathe T, Wirth KE, Holme MP, et al. Botswana's progress toward achieving the 2020 UNAIDS 90-90-90 antiretroviral therapy and virological suppression goals: a population-based survey. *Lancet HIV*. 2016;3(5):e221-230. doi:10.1016/S2352-3018(16)00037-0

7. Dryden-Peterson S, Medhin H, Kebabonye-Pusoentsi M, et al. Correction: Cancer Incidence following Expansion of HIV Treatment in Botswana. *PloS One*. 2015;10(9):e0138742. doi:10.1371/journal.pone.0138742

8. Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy. *Cancer Cell*. 2018;33(4):570-580. doi:10.1016/j.ccell.2018.03.015

9. Klein C, Gonzalez D, Samwel K, et al. Relationship between the Cervical Microbiome, HIV Status, and Precancerous Lesions. *mBio*. 2019;10(1). doi:10.1128/mBio.02785-18

10. Bombardelli C, Ayuso JH, Pelayo RG. A framework for human microbiome research. *Nature*. 2012;486(7402):215-221. doi:10.1038/nature11209

11. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10(10):996-998. doi:10.1038/nmeth.2604

12. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6). doi:10.1186/gb-2011-12-6-r60

13. Kyrgiou M, Mitra A, Moscicki A-B. Does the vaginal microbiota play a role in the development of cervical cancer? *Transl Res J Lab Clin Med*. 2017;179:168-182. doi:10.1016/j.trsl.2016.07.004

14. Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4680-4687. doi:10.1073/pnas.1002611107

15. Brotman RM, Shardell MD, Gajer P, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis*. 2014;210(11):1723-1733. doi:10.1093/infdis/jiu330

16. van de Wijgert JHHM, Borgdorff H, Verhelst R, et al. The vaginal microbiota: what have we learned after a decade of molecular characterization? *PloS One*. 2014;9(8):e105998. doi:10.1371/journal.pone.0105998

17. King CC, Jamieson DJ, Wiener J, et al. Bacterial vaginosis and the natural history of human papillomavirus. *Infect Dis Obstet Gynecol*. 2011;2011:319460. doi:10.1155/2011/319460

18. Wheeler S, Pryor K, Antczak B, Truong T, Murtha A, Seed P. The relationship of cervical microbiota diversity with race and disparities in preterm birth. *J Neonatal-Perinat Med*. 2018;11(3):305-310. doi:10.3233/NPM-17111

19. Kacerovsky M, Pliskova L, Bolehovska R, et al. Lactobacilli-dominated cervical microbiota in women with preterm prelabor rupture of membranes. *Pediatr Res*. December 2019. doi:10.1038/s41390-019-0692-1

20. Kacerovsky M, Vrbacky F, Kutova R, et al. Cervical microbiota in women with preterm prelabor rupture of membranes. *PloS One*. 2015;10(5):e0126884. doi:10.1371/journal.pone.0126884

21. Wang Y, Zhang Y, Zhang Q, Chen H, Feng Y. Characterization of pelvic and cervical microbiotas from patients with pelvic inflammatory disease. *J Med Microbiol*. 2018;67(10):1519-1526. doi:10.1099/jmm.0.000821

22. Marullo R, Werner E, Zhang H, Chen GZ, Shin DM, Doetsch PW. HPV16 E6 and E7 proteins induce a chronic oxidative stress response via NOX2 that causes genomic instability and increased susceptibility to DNA damage in head and neck cancer cells. *Carcinogenesis*. 2015;36(11):1397-1406. doi:10.1093/carcin/bgv126

23. Mitra A, MacIntyre DA, Lee YS, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci Rep*. 2015;5:16865. doi:10.1038/srep16865

24. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet Lond Engl*. 2001;357(9255):539-545. doi:10.1016/S0140-6736(00)04046-0

25. Balkwill FR, Mantovani A. Cancer-related inflammation: common themes and therapeutic opportunities. *Semin Cancer Biol*. 2012;22(1):33-40. doi:10.1016/j.semcancer.2011.12.005

26. van de Wijgert JHHM, Gill AC, Chikandiwa A, et al. Human papillomavirus infection and cervical dysplasia in HIV-positive women: potential role of the vaginal microbiota. *AIDS*. 2020;34(1):115–125. doi:10.1097/QAD.0000000000002381

27. Karim R, Tummers B, Meyers C, et al. Human papillomavirus (HPV) upregulates the cellular deubiquitinase UCHL1 to suppress the keratinocyte's innate immune response. *PLoS Pathog*. 2013;9(5):e1003384. doi:10.1371/journal.ppat.1003384

28. Ilhan ZE, Łaniewski P, Thomas N, Roe DJ, Chase DM, Herbst-Kralovetz MM. Deciphering the complex interplay between microbiota, HPV, inflammation and cancer through cervicovaginal metabolic profiling. *EBioMedicine*. 2019;44:675-690. doi:10.1016/j.ebiom.2019.04.028

29. Brooks JP, Edwards DJ, Blithe DL, et al. Effects of combined oral contraceptives, depot medroxyprogesterone acetate and the levonorgestrel-releasing intrauterine system on the vaginal microbiome. *Contraception*. 2017;95(4):405-413. doi:10.1016/j.contraception.2016.11.006

30. van de Wijgert JHHM, Jespers V. The global health impact of vaginal dysbiosis. *Res Microbiol*. 2017;168(9-10):859-864. doi:10.1016/j.resmic.2017.02.003

31. Borgdorff H, Gautam R, Armstrong SD, et al. Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. *Mucosal Immunol*. 2016;9(3):621-633. doi:10.1038/mi.2015.86

32. Wu S, Rhee K-J, Albesiano E, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med*. 2009;15(9):1016-1022. doi:10.1038/nm.2015

33. Arthur JC, Perez-Chanona E, Mühlbauer M, et al. Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science*. 2012;338(6103):120-123. doi:10.1126/science.1224820

34. Bullman S, Pedamallu CS, Sicinska E, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science*. 2017;358(6369):1443-1448. doi:10.1126/science.aal5240

35. Yang Y, Weng W, Peng J, et al. Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor-κB, and Up-regulating Expression of MicroRNA-21. *Gastroenterology*. 2017;152(4):851-866.e24. doi:10.1053/j.gastro.2016.11.018

36. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe*. 2013;14(2):195-206. doi:10.1016/j.chom.2013.07.012

Table Legends

Table 1. Clinico-pathological features of patients in Botswana with cervical dysplasia or cervical cancer

Table 2. Selected characteristics of patients in Botswana with cervical dysplasia vs cervical cancer

Figure Legends

Figure 1 Cervical microbiota of cervical dysplasia and cervical cancer in patients with and without HIV. A) Overall alpha diversity, as assessed by Shannon diversity in HIV-positive and -negative cervical dysplasia and cervical cancer patients. B) Beta diversity, as assessed by Bray-Curtis unweighted UniFrac in HIV-positive vs -negative patients. C) Stacked bar plot of the top 10 most abundant genus-level bacteria in HIV-positive vs -negative patients. Each bar represents a single patient and is labeled with the subject's age. D,E) LEfSe identified the most differentially abundant taxa between HIV-positive and -negative patients. D) Cladogram representation of the significantly different taxa features, from phylum (inner circle) to genus (outer circle). E) Histogram showing the LDA scores of genera that were differentially abundant between the 2 groups. The LEfSe was restricted to $p<0.05$ for the class and subclass analysis and a minimum LDA score of 2.0.

Figure 2 Cervical microbiota in cervical cancer patients is statistically significantly different from that in cervical dysplasia patients. A,B) Overall alpha diversity, as assessed by Shannon diversity in cervical dysplasia and cervical cancer patients. C,D) Beta diversity, as assessed by Bray-Curtis weighted UniFrac in cervical dysplasia vs cervical cancer patients. E) Stacked bar plot of the top 10 most abundant genus-level bacteria in cervical dysplasia patients vs cervical cancer patients. Each bar represents a single participant and is labeled with the subject's age. D,E) LEfSe identified the most differentially abundant taxa in cervical dysplasia and cervical cancer patients. D) Cladogram representation of the significantly different taxa features, from phylum (inner circle) to genus (outer circle). E) Histogram showing the LDA

446 scores of genera that were differentially abundant between the 2 groups. LEfSe was restricted to

447 $p<0.05$ for the class and subclass analysis and a minimum LDA score of 2.0.

Table 1 Clinico-pathological features of 31 patients in Botswana with cervical dysplasia or cervical cancer

Feature	Result
Type of cervical lesion, n	
CIN stage I	0
CIN stage II	3
CIN stage III	18
Cervical cancer	10
HIV status, %	
Positive	77
Negative	23
Smoking status, %	
Smoker	7
Non-Smoker	94

CIN, cervical intraepithelial neoplasia.

Table 2 Selected characteristics of 31 patients in Botswana with cervical dysplasia vs cervical cancer

Characteristic	Dysplasia (n=21)	Cancer (n=10)	P value*
Mean age (SD), years	41.8 (7.5)	50.7 (12)	0.1
Mean BMI (SD), kg/m^2	26.3 (6.4)	30.0 (7.2)	0.2
HIV status, %			
Positive	81	70	0.5
Negative	19	30	0.5
Smoking status, %			
Smoker	10	0	0.3
Non-Smoker	91	100	0.3

*P values were based on a t-test (continuous variables) or z-test (proportions). All tests were 2-sided.

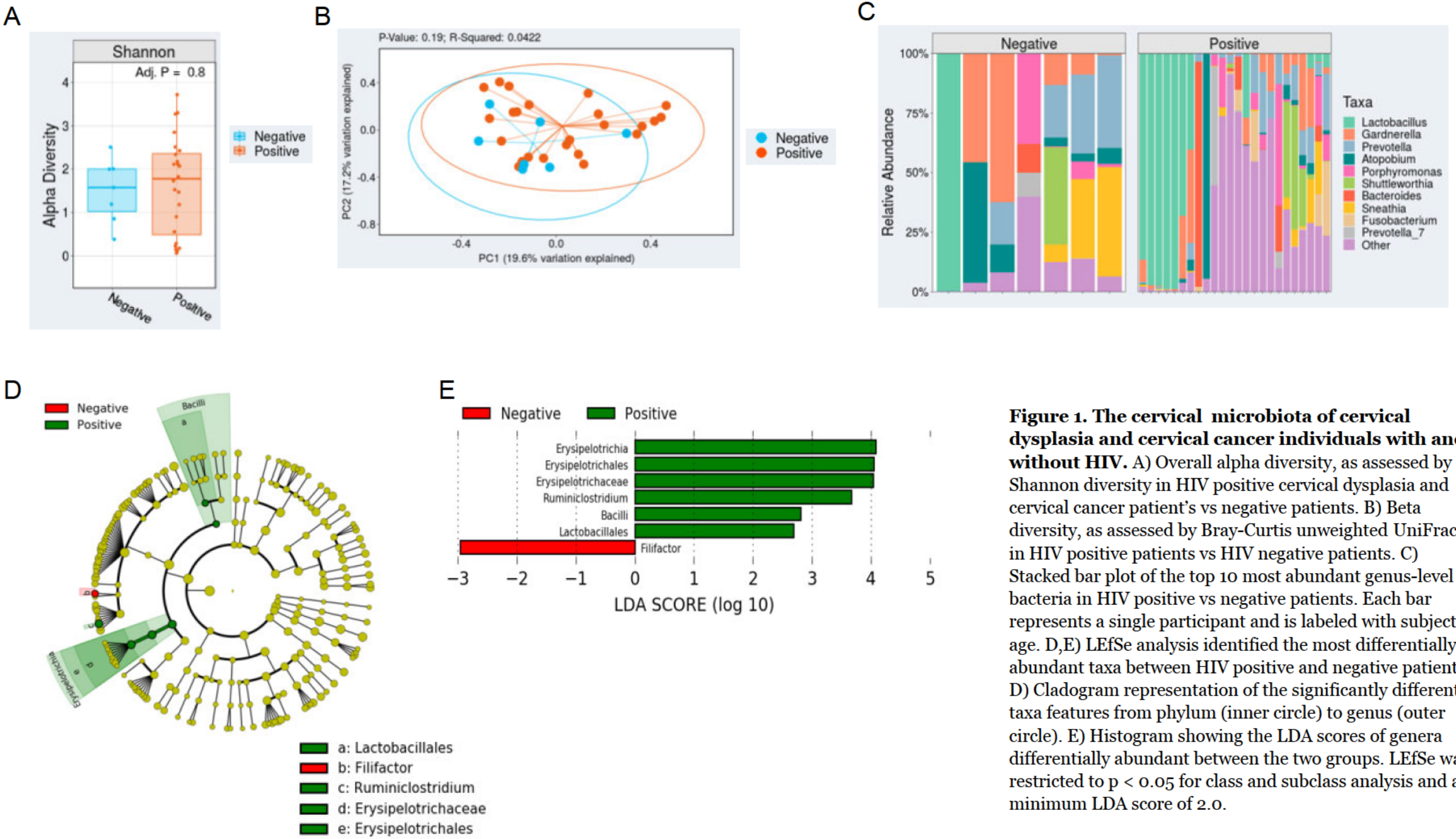


Figure 1. The cervical microbiota of cervical dysplasia and cervical cancer individuals with and without HIV. A) Overall alpha diversity, as assessed by Shannon diversity in HIV positive cervical dysplasia and cervical cancer patient's vs negative patients. B) Beta diversity, as assessed by Bray-Curtis unweighted UniFrac in HIV positive patients vs HIV negative patients. C) Stacked bar plot of the top 10 most abundant genus-level bacteria in HIV positive vs negative patients. Each bar represents a single participant and is labeled with subject age. D,E) LEfSe analysis identified the most differentially abundant taxa between HIV positive and negative patients. D) Cladogram representation of the significantly different taxa features from phylum (inner circle) to genus (outer circle). E) Histogram showing the LDA scores of genera differentially abundant between the two groups. LEfSe was restricted to $p < 0.05$ for class and subclass analysis and a minimum LDA score of 2.0.

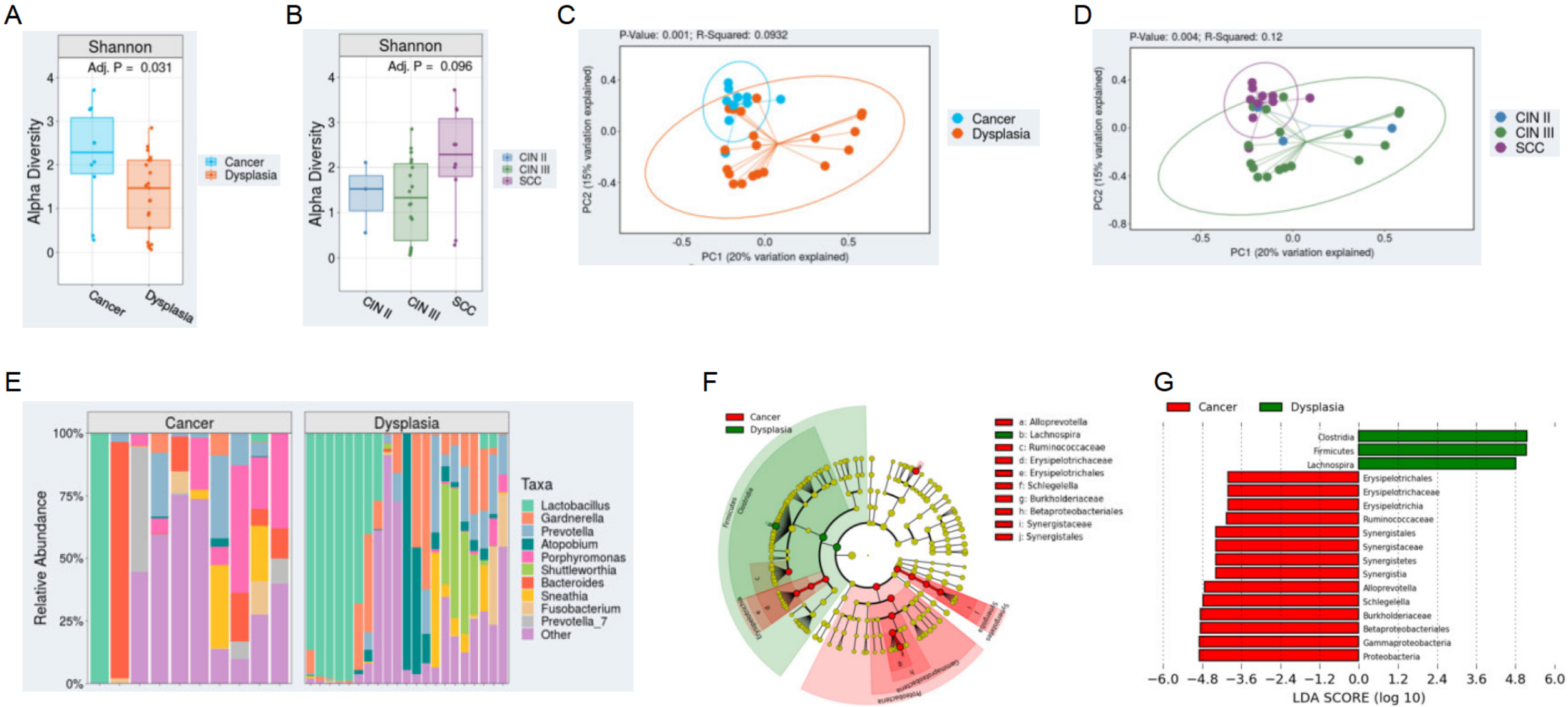


Figure 2. The cervical microbiota with cervical cancer is statistically significantly different individuals with cervical dysplasia. A,B) Overall alpha diversity, as assessed by Shannon diversity in cervical dysplasia and cervical cancer patients. C,D) Beta diversity, as assessed by Bray-Curtis weighted UniFrac in Cervical dysplasia vs cervical cancer patients. E) Stacked bar plot of the top 10 most abundant genus-level bacteria in cervical dysplasia patients vs cervical cancer patients. Each bar represents a single participant and is labeled with subject age. D,E) LEfSe analysis identified the most differentially abundant taxa between cervical dysplasia and cervical cancer patients. D) Cladogram representation of the significantly different taxa features from phylum (inner circle) to genus (outer circle). E) Histogram showing the LDA scores of genera differentially abundant between the two groups. LEfSe was restricted to $p < 0.05$ for class and subclass analysis and a minimum LDA score of 2.0.

Date : 2/13/2020 8:21:24 AM
From : "Ajami,Nadim J"
To : "Javornik Cregeen, Sara Joan" [redacted]
Cc : " [redacted] **, "Wong, Matthew C."** [redacted]
Subject : Re: [EXT] bioRxiv -- Manuscript Closed

Yeah. I was testing the submission process and never finish it. I guess this is what they closed. Sorry for the false alarm. I'm predisposed with biorxiv.

Sent from my iPhone

> On Feb 13, 2020, at 8:11 AM, Javornik Cregeen, Sara Joan [redacted] wrote:
>
> That's not our manuscript though.
>
> This is us:
> BIORXIV/2020/939207
> Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019
> Matthew C Wong, Sara J Javornik Cregeen, Nadim J Ajami, and Joseph F Petrosino
>
> ---
>
> **From:** "Ajami,Nadim J" <NAjami@mdanderson.org>
> **Date:** Thursday, February 13, 2020 at 7:57 AM
> **To:** "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
> **Subject:** Fwd: [EXT] bioRxiv -- Manuscript Closed
>
> Sent from my iPhone
>
> Begin forwarded message:
>
> > **From:** "biorxiv@cshlbp.org" <biorxiv@cshlbp.org>
> > **Date:** February 13, 2020 at 12:09:50 AM CST
> > **To:** "Ajami,Nadim J" <NAjami@mdanderson.org>
> > **Subject: [EXT] bioRxiv -- Manuscript Closed**
> >
> > WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

MS ID#: BIORXIV/2020/925941
MS TITLE: nCoV Spike Protein S1 CTD subdomain Shares High Amino Acid Identity With a Coronavirus Recovered from a Pangolin Viral Metagenomic Dataset

Dear Nadim Ajami;

The above manuscript has been closed.

The bioRxiv team

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 1/30/2020 10:43:29 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Lloyd, Richard E." [redacted]
Cc : "Wong, Matthew C." [redacted]
Subject : Re: [EXT] Re: nCoV analysis

Thank you, Rick!
We will post this asap.
Nadim

From: "Lloyd, Richard E." [redacted]
Date: Thursday, January 30, 2020 at 10:40 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>
Cc: "Wong, Matthew C." [redacted]
Subject: [EXT] Re: nCoV analysis

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi guys,
OK just got a look at this and Matt stopped by my office. I think this looks really nice and is a good way to go. You may want to include a reference for VirMAP ("VirMAP (Nature Commun. 9:3205). Go for it.
Rick

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, January 30, 2020 at 9:52 AM
To: Rick Lloyd [redacted]
Cc: "Wong, Matthew C." [redacted]
Subject: Re: nCoV analysis

Updated text:

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. We used VirMAP to recover potential viral genomes and compare recovered coronaviruses to the outbreak strain. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – Figure 1A. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between the two at amino acid residues 350 to 550. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298), we were able to reconstruct a coronavirus genome (approximately 84% complete from samples SRR10168377 and SRR10168378) that shared 97% amino acid identity across the same RBD segment – Figure 1B. This result indicates a potential recombination event for nCoV-2019.

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, January 30, 2020 at 9:20 AM

To: [redacted]
Cc: [redacted]
Subject: nCoV analysis

Hi Rick,
Hope you are well!

Matt and I got together last night to review his analysis on the recent nCoV-2019 genome. We came up with the following statement summarizing his findings and before posting to Virological.org we wanted to run it by you. Figures attached. Let us know what you think.

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. We used VirMAP to recover potential viral genomes and compare recovered coronaviruses to the outbreak strain. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – **Figure 1A**. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between amino acid residues 350 to 550. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298), we were able to reconstruct a coronavirus genome (approximately 84% complete from sample SRR10168377) that shared 97% amino acid identity across the same RBD genome – **Figure 1B**. This result indicates a potential recombination event for nCoV-2019.

VirMAP-Pangolin CoV genome reconstruction: *google drive link*

Best,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 1/30/2020 10:53:00 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Wong, Matthew C." [redacted]
Subject : Re: [EXT] Re: nCoV analysis
Attachment : nCoV-2019.docx;

From: "Wong, Matthew C." [redacted]
Date: Thursday, January 30, 2020 at 10:17 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Lloyd, Richard E." [redacted]
Subject: [EXT] Re: nCoV analysis

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – Figure 1A. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between the two atamino acid residues 350 to 550. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298) we used VirMAP to reconstruct a coronavirus genome (approximately 84% complete from samples SRR10168377 and SRR10168378) that shared 97% amino acid identity across the same RBD segment – Figure 1B. This result indicates a potential recombination event for nCoV-2019.

nCoV Spike Protein Receptor Binding Domain Shares High Amino Acid Identity With a Coronavirus Recovered from a Pangolin Viral Metagenomic Dataset

An outbreak of respiratory illness caused by a novel coronavirus (nCoV-2019, NC_045512.2) first identified in Wuhan China has resulted in over seven thousand confirmed cases. So far, the nCoV-2019 has been reported to share 96% sequence identity to the RaTG13 genome (EPI_ISL_402131) – Figure 1A. However, the S1 Receptor Binding Domain (RBD) of the nCoV-2019 genome was noticeably divergent between the two at amino acid residues 350 to 550. We aimed to identity coronaviruses related to nCoV-2019 in viral metagenomics datasets available in the public domain. In a recently published dataset describing viral diversity in Malayan pangolins (doi:10.3390/v11110979, PRJNA573298) we used VirMAP (doi.org/10.1038/s41467-018-05658-8) to reconstruct a coronavirus genome (approximately 84% complete from samples SRR10168377 and SRR10168378) that shared 97% amino acid identity across the same RBD segment – Figure 1B. This result indicates a potential recombination event for nCoV-2019.

NC_045512.2
https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2

EPI_ISL_402131
https://gisaid.org/CoV2020

Malayan Pangolins Paper
https://doi:10.3390/v11110979

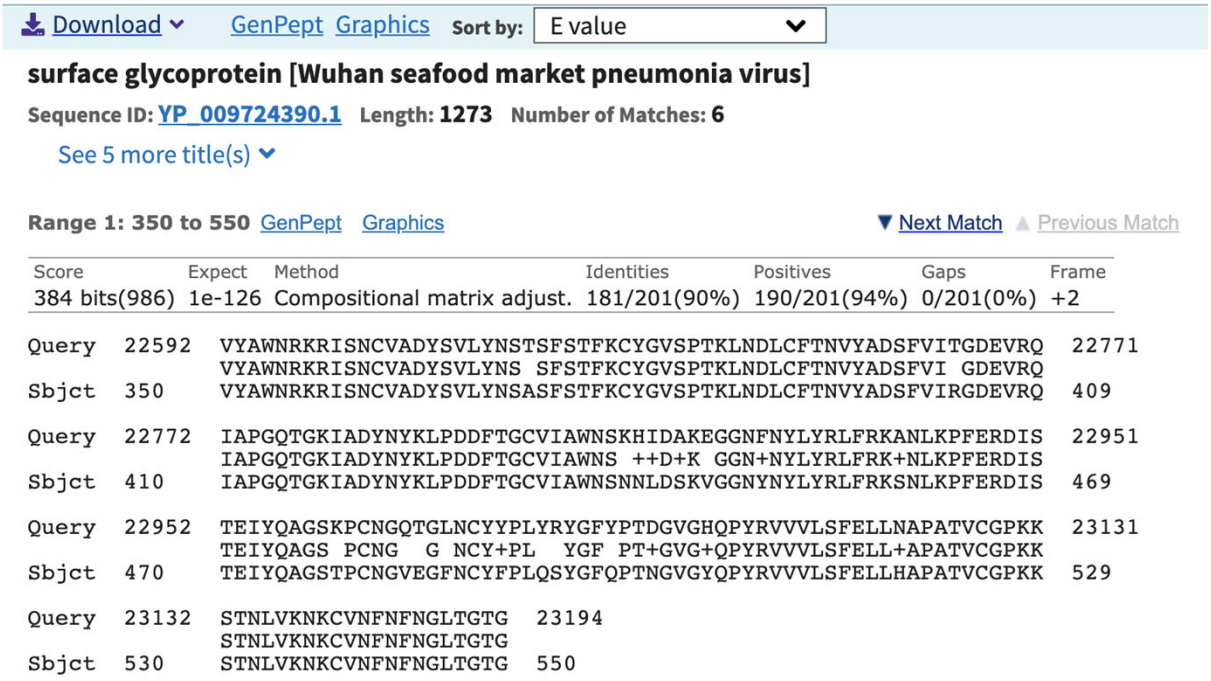
Malayan Pangolins Dataset
https://www.ncbi.nlm.nih.gov/bioproject/573298

VirMAP paper
https://doi.org/10.1038/s41467-018-05658-8

VirMAP – Pangolin Coronavirus fasta:

Figure 1

A. nCoV-2019 vs. RaTG13



Download GenPept Graphics Sort by: E value

surface glycoprotein [Wuhan seafood market pneumonia virus]

Sequence ID: YP_009724390.1 **Length:** 1273 **Number of Matches:** 6

See 5 more title(s)

Range 1: 350 to 550 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
384 bits(986)	1e-126	Compositional matrix adjust.	181/201(90%)	190/201(94%)	0/201(0%)	+2

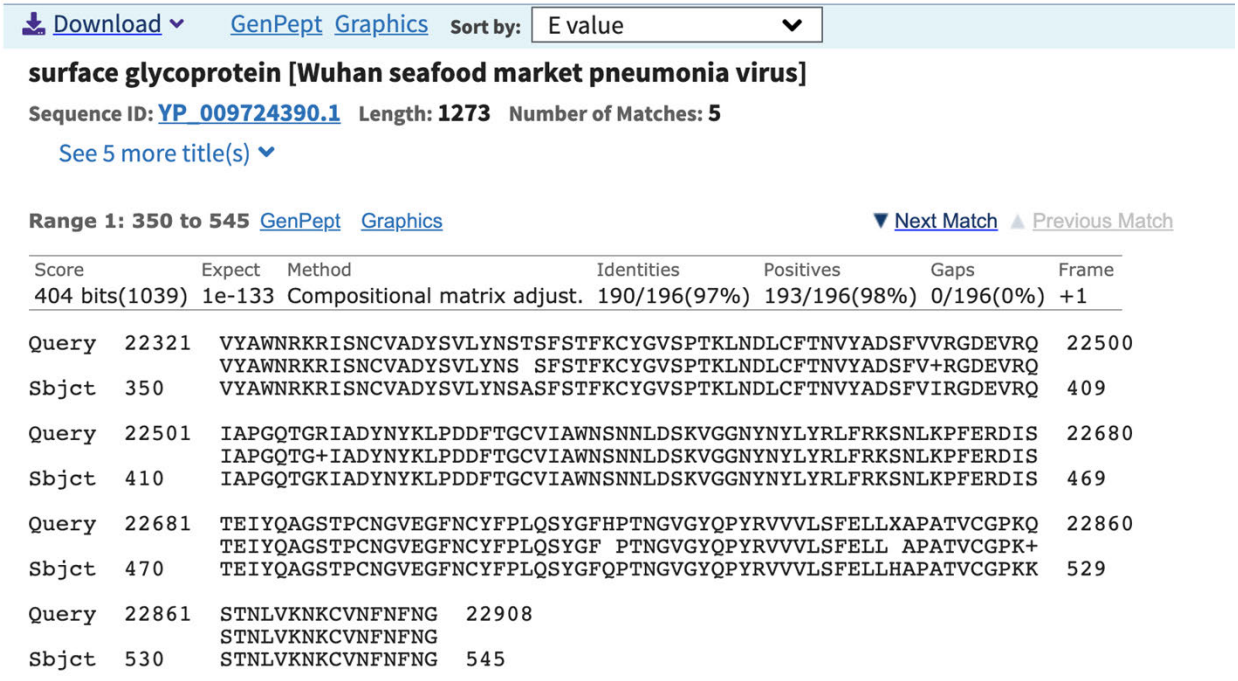
```
Query  22592  VYAWNRKRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVRQ  22771
              VYAWNRKRISNCVADYSVLYNS SFSTFKCYGVSPTKLNDLCFTNVYADSFVI GDEVRQ
Sbjct  350    VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQ  409

Query  22772  IAPGQTGKIADYNYKLPDDFTGCVIAWNSKHIDAKEGGNFNYLYRLFRKANLKPFERDIS  22951
              IAPGQTGKIADYNYKLPDDFTGCVIAWNS ++D+K GGN+NYLYRLFRK+NLKPFERDIS
Sbjct  410    IAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS  469

Query  22952  TEIYQAGSKPCNGQTGLNCYYPLYRYGFYPTDGVGHQPYRVVVLSFELLNAPATVCGPKK  23131
              TEIYQAGS PCNG  G NCY+PL  YGF PT+GVG+QPYRVVVLSFELL+APATVCGPKK
Sbjct  470    TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKK  529

Query  23132  STNLVKNKCVNFNFNGLTGTG  23194
              STNLVKNKCVNFNFNGLTGTG
Sbjct  530    STNLVKNKCVNFNFNGLTGTG  550
```

B. nCoV-2019 vs. Pangolin CoV



Download GenPept Graphics Sort by: E value

surface glycoprotein [Wuhan seafood market pneumonia virus]

Sequence ID: YP_009724390.1 **Length:** 1273 **Number of Matches:** 5

See 5 more title(s)

Range 1: 350 to 545 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
404 bits(1039)	1e-133	Compositional matrix adjust.	190/196(97%)	193/196(98%)	0/196(0%)	+1

```
Query  22321  VYAWNRKRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVRGDEVRQ  22500
              VYAWNRKRISNCVADYSVLYNS SFSTFKCYGVSPTKLNDLCFTNVYADSFV+RGDEVRQ
Sbjct  350    VYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQ  409

Query  22501  IAPGQTGRIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS  22680
              IAPGQTG+IADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS
Sbjct  410    IAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIS  469

Query  22681  TEIYQAGSTPCNGVEGFNCYFPLQSYGFHPTNGVGYQPYRVVVLSFELLXAPATVCGPKQ  22860
              TEIYQAGSTPCNGVEGFNCYFPLQSYGF PTNGVGYQPYRVVVLSFELL APATVCGPK+
Sbjct  470    TEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKK  529

Query  22861  STNLVKNKCVNFNFNG  22908
              STNLVKNKCVNFNFNG
Sbjct  530    STNLVKNKCVNFNFNG  545
```

Date : 4/16/2020 9:50:46 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Samantha Coy" [redacted]**, "Wilhelm, Steven W"** [redacted]
Cc : " [redacted] **"Gann, Eric"** [redacted]
Subject : Re: [EXT] Fwd: Frontiers: Congratulations! Your manuscript is accepted - 532536

Awesome news, Sam!
All the best,
Nadim

From: Samantha Coy [redacted]
Date: Thursday, April 16, 2020 at 8:24 AM
To: "Wilhelm, Steven W" [redacted]
Cc: [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, [redacted] "Gann, Eric" [redacted]
Subject: [EXT] Fwd: Frontiers: Congratulations! Your manuscript is accepted - 532536

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi everyone,

I think you all have received notification that our manuscript was accepted for publication, but in any case, I wanted to let everyone know as a group and pass on my gratefulness to each of you! Your contributions are much appreciated, and it feels so good to have this finished!

Hope you are all doing well with everything going on.

All the very best,

Samantha

---------- Forwarded message ---------
From: **Frontiers Microbiology Editorial Office** <microbiology.editorial.office@frontiersin.org>
Date: Thu, Apr 16, 2020 at 4:49 AM
Subject: Frontiers: Congratulations! Your manuscript is accepted - 532536

To: [redacted]

Dear Dr Coy,

Frontiers Microbiology Editorial Office has sent you a message. Please click 'Reply' to send a direct response

I am pleased to inform you that your manuscript SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems has been approved for production and accepted for publication in Frontiers in Microbiology, section Virology.
Your manuscript is currently being prepared for publication. The provisional version of the abstract or introductory section is currently available online. Please do not communicate any changes at this stage. You will be contacted as soon as the author proofs are ready for your revisions.

Manuscript title: SMRT sequencing of Paramecium bursaria Chlorella Virus-1 reveals diverse methylation stability in adenines targeted by restriction modification systems
Journal: Frontiers in Microbiology, section Virology
Article type: Original Research
Authors: Samantha R Coy, Eric Robert Gann, Spiridon E Papoulis, Michael Holder, Nadim Ajami, Joseph Petrosino, Erik Zinser, James L Van Etten, Steven W Wilhelm
Manuscript ID: 532536
Edited by: Andrew S Lang

You can click here to access the final review reports and manuscript:
http://www.frontiersin.org/Review/EnterReviewForum.aspx?activationno=80dfc921-8a82-4e86-a249-6bc5e33b7d34

As an author, it is important that you maintain your Frontiers research network (Loop) profile up to date, as your publication will be linked to your profile allowing you and your publications to be more discoverable. You can update profile pages (profile pictures, short bio, list of publications) using this link: http://loop.frontiersin.org/people/

Tell us what you think!

At Frontiers we are constantly trying to improve our Collaborative Review process and would like to get your feedback on how we did. Please complete our short 3-minute survey and we will donate $1 to Enfants du Monde, a Swiss non-profit organization:
https://frontiers.qualtrics.com/jfe/form/SV_8q8kYmXRvxBH5at?survey=author&aid=532536&uid=877766

Thank you very much for taking the time to share your thoughts.

Best regards,

Your Frontiers in Microbiology team

Frontiers | Editorial Office - Collaborative Peer Review Team

www.frontiersin.org
Avenue du Tribunal Fédéral 34, 1005 Lausanne, Switzerland
Office T 41 21 510 17 25

For technical issues, please contact our IT Helpdesk (support@frontiersin.org) or visit our Frontiers Help Center (zendesk.frontiersin.org/hc/en-us)

Date : 4/16/2020 7:07:08 AM
From : "Wargo,Jennifer" JWargo@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]
Cc : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org, "Wong, Matthew C." [redacted] **"Ajami,Nadim J" NAjami@mdanderson.org**
Subject : Re: [EXT] Re: MetaPhlan2

Thx Kristi

Sent from my iPhone

On Apr 16, 2020, at 6:57 AM, Hoffman, Kristi Louise [redacted] wrote:

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

This request is in Sara's queue, and she will complete it as soon as her urgent COVID tasks are done. She expects to have it Friday.

Kristi

From: "Khan,Md Abdul Wadud" <MKhan7@mdanderson.org>
Date: Saturday, April 11, 2020 at 9:19 PM
To: "Hoffman, Kristi Louise" [redacted]
Cc: "Wong, Matthew C." [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wargo,Jennifer" <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

Hope you are staying safe and healthy.

Wondering whether you have any update on the metaphlan2?

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Friday, March 27, 2020 10:52 AM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>;

Petrosino, Joseph [redacted]
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud (and team),

The earliest the MetaPhlAn2 request can be completed is the week of April 6th. Let me know if you'd still like us to process the data given that timeframe.

Please note that with regards to Virmap, data processing requests need to go through a project manager and completed according to our queue. While we can expedite requests, especially for *trusted*, long-term collaborators, proper procedures still need to be followed. Circumventing these procedures affects other valued CMMR collaborators and is not taken lightly. I expect this won't be an issue going forward and any requests will go through the proper channels.

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 3:44 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am actually hoping to get the output of MetaPhlan2 by this week but if you can get it done by next week that would be great too.

I already got the output of VirMap. So, no worry on this analysis.

Best

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Wednesday, March 25, 2020 2:47 PM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>

Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted]>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run

them through the MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/18/2020 4:51:31 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Javornik Cregeen, Sara Joan" [redacted]
"Petrosino, Joseph" [redacted] **"Hoffman, Kristi Louise"**
[redacted] **"Wong, Matthew C."** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/19/2020 10:16:34 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted] **"Javornik Cregeen, Sara Joan"** [redacted] **"Petrosino, Joseph"** [redacted] **"Wong, Matthew C."** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted]
"Petrosino, Joseph" [redacted] "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research

Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM

To: "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/19/2020 10:02:45 AM
From : "Hoffman, Kristi Louise" [redacted]
To : "Javornik Cregeen, Sara Joan" [redacted]
"Petrosino, Joseph" [redacted], **"Ajami,Nadim J"**
NAjami@mdanderson.org, "Wong, Matthew C." [redacted]
Subject : RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise <[redacted] Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The ‘benchmarking’ proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn’t qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I’ll tell them it was decided as a no-go. They’ll decide if they want to wait for the installer to be up or move forward with their current results. It’s a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew
C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We’d be happy to assist. However, “help[ing] them benchmark their results” is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted]
"Petrosino, Joseph" [redacted], "Hoffman, Kristi
Louise" [redacted], "Wong, Matthew
C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,

Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly

referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/19/2020 9:02:58 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted] **'Javornik Cregeen, Sara Joan"** [redacted] **"Petrosino, Joseph"** [redacted] **'Wong, Matthew C."** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph [redacted] 'Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph [redacted] "Hoffman, Kristi Louise [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] 'Javornik Cregeen, Sara Joan" [redacted] "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/19/2020 6:03:48 AM
From : "Hoffman, Kristi Louise" [redacted]
To : "Ajami,Nadim J" [redacted]**, "Javornik Cregeen, Sara Joan** [redacted]**"Petrosino, Joseph"** [redacted]**'Wong, Matthew C."** [redacted]
Subject : Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/26/2020 12:45:19 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]**, "Javornik Cregeen, Sara Joan"** [redacted]**, "Petrosino, Joseph"** [redacted]**, "Wong, Matthew C."** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.

The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM

To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted],

"Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] >, "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/26/2020 12:43:04 PM
From : "Hoffman, Kristi Louise" [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org, "Javornik Cregeen, Sara Joan" [redacted] **, "Petrosino, Joseph"** [redacted] **, "Wong, Matthew C."** [redacted]
Subject : RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind “benchmarking” is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn’t run a script.

It’s rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I’m not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The ‘benchmarking’ proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn’t qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I’ll tell them it was decided as a no-go. They’ll decide if they want to wait for the installer to be up or move forward with their current results. It’s a small

dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew
C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted]
"Petrosino, Joseph" [redacted], "Hoffman, Kristi
Louise" [redacted], "Wong, Matthew
C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/26/2020 12:35:59 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]**, "Javornik Cregeen, Sara Joan"** [redacted]**, "Petrosino, Joseph"** [redacted]**, "Wong, Matthew C."** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted] "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew

C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think

that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/26/2020 2:30:04 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Javornik Cregeen, Sara Joan" [redacted], **"Hoffman, Kristi Louise"** [redacted], **"Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]>
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." <[redacted]>
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted]>, "Petrosino, Joseph" [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a

script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To:
>, "Petrosino, Joseph" [redacted], "Wong, Matthew
C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted],
"Petrosino, Joseph" [redacted], "Hoffman, Kristi
Louise" [redacted], "Wong, Matthew
C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino,
Joseph" [redacted], "Hoffman, Kristi
Louise" [redacted], "Wong, Matthew
C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/27/2020 5:16:58 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Javornik Cregeen, Sara Joan" [redacted]
"Hoffman, Kristi Louise" [redacted]**, "Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Thank you, Sara and team.
I'll make sure that everyone is acknowledged appropriately.
Very best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted] Wong, Matthew C [redacted]

Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" <[redacted]>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Ajami,Nadim J" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." <[redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you

have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/27/2020 4:45:17 PM
From : "Javornik Cregeen, Sara Joan" [redacted]
To : "Ajami,Nadim J" [redacted] **, "Hoffman, Kristi Louise"** [redacted] **, "Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted] Wong, Matthew C [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]

Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424

[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" <[redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" <[redacted]>, "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" <[redacted]>, "Wong, Matthew C." <[redacted]>
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are

not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/27/2020 7:30:02 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Javornik Cregeen, Sara Joan" [redacted]
"Hoffman, Kristi Louise" [redacted]**, "Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run
Attachment : VirMAP_Deliverables.docx;

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" [redacted]**, "Hoffman, Kristi Louise"** [redacted]**, "Petrosino, Joseph"** [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" [redacted]**, "Petrosino, Joseph"** [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" [redacted] "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino,

Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release.

Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

VirMAP Deliverables

Final.fa
Assembled genomes in fasta format.

Read Stats
Distribution of raw and trimmed read pairs. VirMAP's default trimming parameters are set to entropy = 0.7, and kmer length = 10.

Called Reads
Number of reads assigned to the virus super-kingdom by VirMAP.

Virome Reads
List of viral taxa and corresponding reads assigned by VirMAP.

Coverage
Genome coverage as determined by the number of reads assigned to each viral taxon over the genome length of the virus identified.

Bit Score (information content)
The sum total of aligned bits per genome calculated at the base level and representing the overall quality of alignment.

Score Ratio
Ratio of the observed bit score to the maximum possible bit score of aligned segments expressed in percentages.

Read Overlap
A measurement for the overlap of viral reads used in each assembly. Values represent the underlying diversity of genomic segments constructed.

Date : 5/28/2020 9:51:30 AM
From : "Javornik Cregeen, Sara Joan" [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org, "Hoffman, Kristi Louise" [redacted] **"Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

No, the results are not compiled. I sent you just the default outputs of a standard Virmap run. The tables aren't actually part of the pipeline, but I can't generate them for you. The Read Stats will probably be different to what is on your list, since we do the trimming prior to the actual Virmap algorithm and I have my own compiler for that.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Wednesday, May 27, 2020 at 7:33 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:

https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,

Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted], Joseph [redacted] Wong, Matthew C [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]

Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise <[redacted]>; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are

not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/28/2020 10:50:20 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Javornik Cregeen, Sara Joan" [redacted], **"Hoffman, Kristi Louise"** [redacted] **"Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Thanks for the quick response, Sara.
Could you clarify if the results can or can't be compiled? Your email says can't but I think you meant can – hopefully, I am right ☐
I'll take whatever you can give me.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 9:51 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

No, the results are not compiled. I sent you just the default outputs of a standard Virmap run. The tables aren't actually part of the pipeline, but I can't generate them for you. The Read Stats will probably be different to what is on your list, since we do the trimming prior to the actual Virmap algorithm and I have my own compiler for that.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Wednesday, May 27, 2020 at 7:33 PM
To: "Javornik Cregeen, Sara Joan" <[redacted]>, "Hoffman, Kristi Louise" <[redacted]>, "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM

To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,

Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph <[redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are

not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/28/2020 4:39:28 PM
From : "Javornik Cregeen, Sara Joan" [redacted]
To : "Ajami,Nadim J" NAjami@mdanderson.org, "Hoffman, Kristi Louise" [redacted] **"Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run
Attachment : CopenhagenVirmapTables.zip;

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Sorry, yes I did mean I CAN generate the tables! Doing too many things at once…

I've attached a zip with all the various tables. It occurred to me while generating these that I didn't ask what type of sample these are, but just assumed they were human. Part of our standard pipeline is the human filtering step that removes host reads – looking at the Read Stats table there aren't very many. I don't know if this means it wasn't a human dataset or that they prefiltered. In any case, if the former is the case and you see an issue with the human filtering step let me know and I'll re-run without it.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, May 28, 2020 at 10:50 AM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks for the quick response, Sara.
Could you clarify if the results can or can't be compiled? Your email says can't but I think you meant can – hopefully, I am right ☐
I'll take whatever you can give me.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 9:51 AM
To: "Ajami,Nadim J" [redacted] "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

No, the results are not compiled. I sent you just the default outputs of a standard Virmap run. The tables aren't actually part of the pipeline, but I can't generate them for you. The Read Stats will probably be different to what is on your list, since we do the trimming prior to the actual Virmap algorithm and I have my own compiler for that.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Wednesday, May 27, 2020 at 7:33 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" <[redacted]>, "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email

address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" [redacted]>, "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." <[redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" <[redacted]>, "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any

further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and

delete all references to it and its contents from your systems.

Date : 5/29/2020 12:34:59 AM
From : "Hoffman, Kristi Louise" [redacted]
To : "Javornik Cregeen, Sara Joan" [redacted]
"Ajami,Nadim J" NAjami@mdanderson.org, "Petrosino, Joseph" [redacted]
Subject : Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Please note that the tables Sara provided are not default outputs of Virmap—they are a product of the CMMR, one typically reserved for fee-paying users and funded grant collaborators. The default outputs of Virmap (both in its published and current forms) were the original files Sara sent. If you share the tables with the Copenhagen group, they should be made aware of this fact.

Best,

Kristi

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 4:39 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Nadim,

Sorry, yes I did mean I CAN generate the tables! Doing too many things at once…

I've attached a zip with all the various tables. It occurred to me while generating these that I didn't ask what type of sample these are, but just assumed they were human. Part of our standard pipeline is the human filtering step that removes host reads – looking at the Read Stats table there aren't very many. I don't know if this means it wasn't a human dataset or that they prefiltered. In any case, if the former is the case and you see an issue with the human filtering step let me know and I'll re-run without it.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, May 28, 2020 at 10:50 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks for the quick response, Sara.
Could you clarify if the results can or can't be compiled? Your email says can't but I think you meant can – hopefully, I am right ☐
I'll take whatever you can give me.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 9:51 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

No, the results are not compiled. I sent you just the default outputs of a standard Virmap run. The tables aren't actually part of the pipeline, but I can't generate them for you. The Read Stats will probably be different to what is on your list, since we do the trimming prior to the actual Virmap algorithm and I have my own compiler for that.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Wednesday, May 27, 2020 at 7:33 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]

Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Ajami,Nadim J" [redacted] "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise <[redacted]>; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted]; Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The ‘benchmarking’ proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn’t qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I’ll tell them it was decided as a no-go. They’ll decide if they want to wait for the installer to be up or move forward with their current results. It’s a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" <[redacted]>
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We’d be happy to assist. However, “help[ing] them benchmark their results” is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,

Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." <[redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted] "Javornik Cregeen, Sara Joan" <[redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any

further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/29/2020 12:57:51 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Javornik Cregeen, Sara Joan" [redacted], **"Hoffman, Kristi Louise"** [redacted] **"Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Thank you, Sara. I'll let you know.
Very best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 4:39 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise'" [redacted] >, "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Sorry, yes I did mean I CAN generate the tables! Doing too many things at once…

I've attached a zip with all the various tables. It occurred to me while generating these that I didn't ask what type of sample these are, but just assumed they were human. Part of our standard pipeline is the human filtering step that removes host reads – looking at the Read Stats table there aren't very many. I don't know if this means it wasn't a human dataset or that they prefiltered. In any case, if the former is the case and you see an issue with the human filtering step let me know and I'll re-run without it.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, May 28, 2020 at 10:50 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks for the quick response, Sara.
Could you clarify if the results can or can't be compiled? Your email says can't but I think you meant can – hopefully, I am right ☐
I'll take whatever you can give me.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 9:51 AM

To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" <[redacted]>, "Petrosino, Joseph" <[redacted]>
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

No, the results are not compiled. I sent you just the default outputs of a standard Virmap run. The tables aren't actually part of the pipeline, but I can't generate them for you. The Read Stats will probably be different to what is on your list, since we do the trimming prior to the actual Virmap algorithm and I have my own compiler for that.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Wednesday, May 27, 2020 at 7:33 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p

2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted] "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" <[redacted]>, "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind “benchmarking” is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn’t run a script.

It’s rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I’m not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030
713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise <[redacted]>; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The ‘benchmarking’ proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn’t qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I’ll tell them it was decided as a no-go. They’ll decide if they want to wait for the installer to be up or move forward with their current results. It’s a small dataset

and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" <[redacted]>, "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information

(PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any

attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/29/2020 12:55:52 AM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]**, "Javornik Cregeen, Sara Joan"** [redacted] **"Petrosino, Joseph"** [redacted]
Subject : Re: [EXT] Re: VirMAP run

Noted, thanks.
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Friday, May 29, 2020 at 12:34 AM
To: "Javornik Cregeen, Sara Joan" [redacted]
"Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Please note that the tables Sara provided are not default outputs of Virmap—they are a product of the CMMR, one typically reserved for fee-paying users and funded grant collaborators. The default outputs of Virmap (both in its published and current forms) were the original files Sara sent. If you share the tables with the Copenhagen group, they should be made aware of this fact.

Best,

Kristi

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 4:39 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Nadim,

Sorry, yes I did mean I CAN generate the tables! Doing too many things at once…

I've attached a zip with all the various tables. It occurred to me while generating these that I didn't ask what type of sample these are, but just assumed they were human. Part of our standard pipeline is the human filtering step that removes host reads – looking at the Read Stats table there aren't very many. I don't know if this means it wasn't a human dataset or that they prefiltered. In any case, if the former is the case and you see an issue with the human filtering step let me know and I'll re-run without it.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Thursday, May 28, 2020 at 10:50 AM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks for the quick response, Sara.
Could you clarify if the results can or can't be compiled? Your email says can't but I think you meant can – hopefully, I am right ☐
I'll take whatever you can give me.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Thursday, May 28, 2020 at 9:51 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

No, the results are not compiled. I sent you just the default outputs of a standard Virmap run. The tables aren't actually part of the pipeline, but I can't generate them for you. The Read Stats will probably be different to what is on your list, since we do the trimming prior to the actual Virmap algorithm and I have my own compiler for that.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Wednesday, May 27, 2020 at 7:33 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Quick question – are the results compiled in any way? I couldn't find summary tables (read stats, called reads, virome reads, coverage, , bit scores, score ratios – early deliverables glossary attached). These were standard deliverables as I recall.
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Wednesday, May 27, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

Here is the link to the Virmap Results, containing VirmapOutputs (per sample directory generated by virmap), VirmapParameters (the files with the settings use), SampleList (list of sample IDs used in the run).

Shareable URL:
https://jplab.s3.amazonaws.com/share/30d/CopenhagenVirmapResults.zip?AWSAccessKeyId=AKIAIHAKQMQQYKNBJKAQ&Expires=1593207504&Signature=tAfb5CCOBH5p2BK%2FkpUxsrvcpZ8k%3D
File size: 6.2G
md5sum: 0105329cb20083ba595abaad508d8df4
Expiration date: Jun 26 2020

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 2:30 PM
To: "Javornik Cregeen, Sara Joan" [redacted] Hoffman, Kristi Louise" [redacted], "Petrosino, Joseph" [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thank you, Sara!
Best,
Nadim

From: "Javornik Cregeen, Sara Joan" [redacted]
Date: Tuesday, May 26, 2020 at 1:52 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Hoffman, Kristi Louise" [redacted] "Petrosino, Joseph" [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

I can have an aws link with the Virmap Outputs ready tomorrow.

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 26, 2020 at 12:45 PM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Thanks, Kristi.

Hi Sara – please let me know what is the ETA.

Very best,
Nadim

From: "Hoffman, Kristi Louise" <[redacted]>
Date: Tuesday, May 26, 2020 at 12:43 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

This is a go, and it's in the queue. Sara, can you provide an ETA for when this will be completed? Thx!

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 26, 2020 12:36 PM
To: Hoffman, Kristi Louise [redacted] Javornik Cregeen, Sara Joan [redacted] Petrosino, Joseph [redacted]; Wong, Matthew C. <[redacted]>
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Wanted to follow-up on this. Could you please let me know if this is a go/no-go?
Thanks,
Nadim

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Tuesday, May 19, 2020 at 10:16 AM
To: "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Petrosino,

Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,
Option #1 is preferred given that option #2 is not possible at this time.
The suggestion of providing them with outputs (option 1) in addition to asking them to run virmap (option 2) was to give them a benchmark.
They haven't asked for this since option 2 is not available yet.
Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 10:02 AM
To: "Javornik Cregeen, Sara Joan" [redacted]>, "Petrosino, Joseph" [redacted] "Ajami,Nadim J" <NAjami@mdanderson.org>, "Wong, Matthew C." [redacted]
Subject: RE: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

To my mind "benchmarking" is an intellectual contribution. Running a script as part of a service (with a fee) may not qualify, but running a script outside of a service or established collaboration certainly does. There would be no data to analyze if someone didn't run a script.

It's rather unfortunate that instructions to successfully run virmap were not vetted and made public at time of publication. If authorship is not on the table, I see two options.

1. We run the script for them and provide outputs—full stop.
2. We provide them with the opportunity to run virmap themselves via Amazon.

I'm not clear what benchmarking you feel is necessary, but if you have concerns about virmap outputs (or Nature Communications has specifically requested further assistance), please let us know so that we may address them.

Best,

Kristi

Kristi L. Hoffman, PhD, MPH
Assistant Professor
Alkek Center for Metagenomics & Microbiome Research
Baylor College of Medicine
Mailstop BCM385, Rm 700B
One Baylor Plaza
Houston, TX 77030

713-798-1424
[redacted]

From: Ajami,Nadim J <NAjami@mdanderson.org>
Sent: Tuesday, May 19, 2020 9:03 AM
To: Hoffman, Kristi Louise [redacted]; Javornik Cregeen, Sara Joan [redacted]; Petrosino, Joseph [redacted] Wong, Matthew C. [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Kristi,

The 'benchmarking' proposal is coming from our side, not theirs. And as it stands, they are not aware of this yet. I had told them authorship would be ideal if the group, including myself ,contributed intellectually to the project AND if got the chance to review all results and final draft. Running a script doesn't qualify as intellectual contribution in my opinion – akin to what CMMR does with MetaPhlAn and HUMAnN.

If this is the only option, I'll tell them it was decided as a no-go. They'll decide if they want to wait for the installer to be up or move forward with their current results. It's a small dataset and it is only DNA data; megahit + blast (standard approach in the VirMAP paper) could get them very close to the finish line.

Let me know.

Thanks,
Nadim

From: "Hoffman, Kristi Louise" [redacted]
Date: Tuesday, May 19, 2020 at 6:03 AM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Javornik Cregeen, Sara Joan" <[redacted]>, "Petrosino, Joseph" [redacted] "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

We'd be happy to assist. However, "help[ing] them benchmark their results" is going to require more than an acknowledgement or reference to the Virmap paper. Sara will be the one to process this dataset, and both she and Joe would deserve authorship for the time, effort, and resources spent to assist the Copenhagen group. If you feel they would be amenable to that, do let us know, and we can start processing their data.

Thanks,

Kristi

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Monday, May 18, 2020 at 4:51 PM
To: "Javornik Cregeen, Sara Joan" [redacted], "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Wong, Matthew C." [redacted]
Subject: Re: [EXT] Re: VirMAP run

Hi Sara,
Great news on getting VirMAP up on Amazon. Once this is up I'll let Nature Comms editor know.
Having the Copenhagen group test VirMAP would be great but I'd argue it will be better if we could help them benchmark their results. I think this would be the best outcome – they'll get data to continue their work (with CPU time, etc.), and then they can run VirMAP and compare results. Let me know your thoughts?
Thanks,
Nadim

From: "Javornik Cregeen, Sara Joan" <[redacted]>
Date: Monday, May 18, 2020 at 4:45 PM
To: "Ajami,Nadim J" <NAjami@mdanderson.org>, "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted] "Wong, Matthew C." [redacted]
Subject: [EXT] Re: VirMAP run

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Nadim,

It seems like Matt will have a working solution for Virmap set up on Amazon pretty soon. The general setup is there, but he needs to write a set of instructions to accompany the release. Our aim is to have it this week or early next week, so we thought that perhaps the Copenhagen team could be a good group to test it out and give feedback on usability.

What do you think?

Thanks,
Sara

From: "Ajami,Nadim J" <NAjami@mdanderson.org>
Date: Friday, May 15, 2020 at 5:08 PM
To: "Petrosino, Joseph" [redacted], "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted] "Wong, Matthew C." [redacted]
Subject: VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information

(PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 4/23/2020 11:33:25 AM
From : "Sims,Travis T." TTSims@mdanderson.org
To : "Mezzari, Melissa" [redacted]**, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Ning,Matthew Stephen" MSNing@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org, "Court,Kyoko" KCourt1@mdanderson.org, "Wu,Xiaogang" XWu10@mdanderson.org, "Delgado Medrano,Andrea Yizel" AYDelgado@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Solley,Travis N" TNSolley@mdanderson.org, "Ahmed-Kaddar,Mustapha" MAhmed10@mdanderson.org, "Petrosino, Joseph"** [redacted] **"Schmeler,Kathleen M" KSchmele@mdanderson.org,** [redacted] **"Nicola** [redacted]
Cc : "Klopp,Ann H" AKlopp@mdanderson.org, "Biegert, Greyson" [redacted]
Subject : Re: [EXT] RES: Manuscript - Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients

Hi All,

In preparation for submission, please let me know if you there is anything you wish to report on you COI disclosure. Thanks!

Best,
Travis

Travis T. Sims, MD, MPH
Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

From: "Mezzari, Melissa" [redacted]
Date: Tuesday, April 14, 2020 at 2:41 PM
To: "Sims,Travis T." <TTSims@mdanderson.org>, "Colbert,Lauren Elizabeth" <LColbert@mdanderson.org>, "Karpinets,Tatiana V" <TVKarpinets@mdanderson.org>, "Ning,Matthew Stephen" <MSNing@mdanderson.org>, "El Alam,Molly B" <MBEl@mdanderson.org>, "Court,Kyoko" <KCourt1@mdanderson.org>, "Wu,Xiaogang" <XWu10@mdanderson.org>, "Delgado Medrano,Andrea Yizel" <AYDelgado@mdanderson.org>, "Ajami,Nadim J" <NAjami@mdanderson.org>, "Solley,Travis N" <TNSolley@mdanderson.org>, "Ahmed-Kaddar,Mustapha" <MAhmed10@mdanderson.org>, "Petrosino,

Joseph" [redacted], "Schmeler,Kathleen M" <KSchmele@mdanderson.org>, [redacted]
Cc: "Klopp,Ann H" <AKlopp@mdanderson.org>, "Biegert, Greyson" <Greyson.Biegert@uth.tmc.edu>
Subject: [EXT] RES: Manuscript - Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Very curious to read this! Klopp's group is on fire with all these new manuscripts! Thank you for including me!

De: Sims,Travis T. <TTSims@mdanderson.org>
Enviada em: Tuesday, April 14, 2020 1:12 PM
Para: Colbert,Lauren Elizabeth <LColbert@mdanderson.org>; Karpinets,Tatiana V <TVKarpinets@mdanderson.org>; Ning,Matthew Stephen <MSNing@mdanderson.org>; El Alam,Molly B <MBEl@mdanderson.org>; Court,Kyoko <KCourt1@mdanderson.org>; Wu,Xiaogang <XWu10@mdanderson.org>; Mezzari, Melissa [redacted] Delgado Medrano,Andrea Yizel <AYDelgado@mdanderson.org>; Ajami,Nadim J <NAjami@mdanderson.org>; Solley,Travis N <TNSolley@mdanderson.org>; Ahmed-Kaddar,Mustapha <MAhmed10@mdanderson.org>; Petrosino, Joseph [redacted] Schmeler,Kathleen M <KSchmele@mdanderson.org>; [redacted]
Cc: Klopp,Ann H <AKlopp@mdanderson.org>; Biegert, Greyson <Greyson.Biegert@uth.tmc.edu>
Assunto: Manuscript - Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients

Hello all,

We have completed the first draft of the manuscript for our project "Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients".

You have been included on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to The *International Journal of Gynecological Cancer* (*IJGC*).

Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.

I am grateful for your time and feedback regarding this project! We hope to submit by 4/24/20.

Best,

Travis

Travis T. Sims, MD, MPH
Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 7/6/2020 8:56:30 AM
From : "Diaz,Christine M" CMDiaz1@mdanderson.org
To : "Risteski,Hristijan" HRisteski@mdanderson.org
Cc : "Wargo,Jennifer" JWargo@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org
Subject : RE: 2nd Request - FW: Need Sub Awards input Due July 6 - FW: R01 PQ10 Progress Report

Ok, I'll send out an invite.

Thank You!

Christine Diaz
WK: 713-745-3225 MDA Cell: 713-598-7411
cmdiaz1@mdanderson.org

From: Risteski,Hristijan <HRisteski@mdanderson.org>
Sent: Monday, July 6, 2020 8:45 AM
To: Diaz,Christine M <CMDiaz1@mdanderson.org>
Cc: Wargo,Jennifer <JWargo@mdanderson.org>
Subject: RE: 2nd Request - FW: Need Sub Awards input Due July 6 - FW: R01 PQ10 Progress Report

If go with NCE, the report will be due next year, same time. 11:00 am is good for me…
Kiko

From: Diaz,Christine M
Sent: Monday, July 06, 2020 8:43 AM
To: Risteski,Hristijan <HRisteski@mdanderson.org>
Cc: Wargo,Jennifer <JWargo@mdanderson.org>; Diaz,Christine M <CMDiaz1@mdanderson.org>
Subject: RE: 2nd Request - FW: Need Sub Awards input Due July 6 - FW: R01 PQ10 Progress Report

Hi

Would this effect the report? I can be open after 11:00 am.

Thank You!

Christine Diaz
WK: 713-745-3225 MDA Cell: 713-598-7411
cmdiaz1@mdanderson.org

From: Risteski,Hristijan <HRisteski@mdanderson.org>
Sent: Sunday, July 5, 2020 8:55 PM

To: Diaz,Christine M <CMDiaz1@mdanderson.org>
Subject: RE: 2nd Request - FW: Need Sub Awards input Due July 6 - FW: R01 PQ10 Progress Report

Hey Christine,

I hope you had a nice 4th of July meeting.

I've been thinking about this project, and I would like to touch base with you tomorrow morning, to see if better option would be to ask for a NCE for Year 2 of the grant.

Many thanks,
Kiko

From: Diaz,Christine M
Sent: Tuesday, June 30, 2020 5:11 PM
To: Hu, Jianhua [redacted]
Cc: Wargo,Jennifer <JWargo@mdanderson.org>; Risteski,Hristijan <HRisteski@mdanderson.org>; Diaz,Christine M <CMDiaz1@mdanderson.org>
Subject: 2nd Request - FW: Need Sub Awards input Due July 6 - FW: R01 PQ10 Progress Report
Importance: High

If you can please review the attached and provide your updates asap. Thank you.

Thank You!

Christine Diaz
WK: 713-745-3225 MDA Cell: 713-598-7411
cmdiaz1@mdanderson.org

From: Diaz,Christine M
Sent: Thursday, June 25, 2020 10:33 AM
To: Hu, Jianhua [redacted]
Cc: Hristijan Risteski (HRisteski@mdanderson.org) <HRisteski@mdanderson.org>; Jennifer Wargo (JWargo@mdanderson.org) <JWargo@mdanderson.org>
Subject: Need Sub Awards input Due July 6 - FW: R01 PQ10 Progress Report
Importance: High

Greetings,

We are preparing the R01 PQ10 Progress Report, can you please provide us by COB on Monday, July 6, the following:

- All personnel report (template attached)
- Other Support for the Key Personnel
- Technical/Narrative report ½ to 1 page, and
- Estimated carryforward balance

I have included the original Application and a copy of last year's report for your reference. If you have any questions, please let us know.

Thank You!

Christine Diaz
WK: 713-745-3225 MDA Cell: 713-598-7411
cmdiaz1@mdanderson.org

From: Risteski,Hristijan <HRisteski@mdanderson.org>
Sent: Tuesday, June 23, 2020 9:27 AM
To: Diaz,Christine M <CMDiaz1@mdanderson.org>
Cc: Wargo,Jennifer <JWargo@mdanderson.org>; Ajami,Nadim J <NAjami@mdanderson.org>
Subject: RE: FW: R01 PQ10 Progress Report

From each of the sub we will need the following sections:

- All personnel report (template attached)
- Other Support for the Key Personnel
- Technical/Narrative report ½ to 1 page, and
- Estimated carryforward balance

These are the subs and key personnel that we had last year:

- Columbia University: Jianhua Hu
- BCM: Joseph Petrosino

Since Reetakshi collected the information last year, I don't have contact information from any of the subs, let me know if you have it.

Please let me know if you have any questions,
Kiko

Date : 4/16/2020 5:42:39 AM
From : "Wargo,Jennifer" JWargo@mdanderson.org
To : "Sims,Travis T." TTSims@mdanderson.org
Cc : "Sastry,Jagannadha K" jsastry@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Schmeler,Kathleen M" KSchmele@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Chapman,Bhavana S" BSChapman@mdanderson.org, "Mezzari, Melissa" [redacted] "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org
Subject : Re: Manuscript - Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation

Nice work!
I will review today and be ready for a discussion tomorrow
Jen

Sent from my iPhone

> On Apr 15, 2020, at 7:20 PM, Sims,Travis T. <TTSims@mdanderson.org> wrote:
>
> Hello all,
> We have completed the first draft of the manuscript for our project "Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation".
>
> Mrs. El Alam, Dr. Colbert, Dr. Klopp and I have included you on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to *Nature Medicine*.
>
> Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.
>
> I am grateful for your time and feedback regarding this project! We hope to submit by 5/1/20.
> Best,
> Travis
>
> Travis T. Sims, MD, MPH
> Fellow
> *Department of Gynecologic Oncology & Reproductive Medicine*
> The University of Texas MD Anderson Cancer Center
> ttsims@mdanderson.org

C [redacted]
T 346-315-9781
P 713-404-6828

<Manuscript - Gut microbiome diversity as an independent predictor of survival in cervical cancer patients receiving chemoradiation V1.docx.awsec>
<Table 1. Gut Microbiome Univariate and Multivariate Analysis RFS 4-15-20.docx.awsec>
<Table 2. Gut Microbiome Univariate and Multivariate Analysis OS 4-15-20.docx.awsec>
<Figures V1 - Gut microbiome diversity an independent predictor of cervical cancer 4-15-2020.pptx.awsec>
<Supplemental Table 1. Baseline diversity vs. demographics 4-15-20.docx.awsec>
<Supplemental Table 2. Gut Microbiome 4-15-20.docx.awsec>
<Supplemental Table 3. Gut Microbiome Univariate All Alpha Diversity Time Points RFS 4-15-20.docx.awsec>
<Supplemental Table 4. Gut Microbiome Univariate All Alpha Diversity Time Points OS 4-15-20.docx.awsec>

Date : 4/19/2020 5:01:21 PM
From : "Schmeler,Kathleen M" KSchmele@mdanderson.org
To : "Sims,Travis T." TTSims@mdanderson.org, "Sastry,Jagannadha K" jsastry@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "Chapman,Bhavana S" BSChapman@mdanderson.org, "Sastry,Jagannadha K" jsastry@mdanderson.org, "Mezzari, Melissa" [redacted] "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "[redacted]
Cc : "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org
Subject : Re: Manuscript - Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation
Attachment : Manuscript - Gut microbiome diversity as an independent predictor of survival in cervical cancer patients receiving chemoradiation V1.KMS.docx;

Hi. See attached comments – thanks for including me!

Kathleen

--
Kathleen M. Schmeler, MD
Professor
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center

Phone: 713-745-3518
Fax: 713-792-7586

Mailing Address:
Unit 1362
PO Box 301439
Houston, TX 77230-1429

From: "Sims,Travis T." <TTSims@mdanderson.org>
Date: Wednesday, April 15, 2020 at 7:20 PM
To: Jagannadha Sastry <jsastry@mdanderson.org>, "Karpinets,Tatiana V" <TVKarpinets@mdanderson.org>, "Lin,Lilie L" <LLLin@mdanderson.org>, Lois Ramondetta <lramonde@mdanderson.org>, Anuja Jhingran <ajhingra@mdanderson.org>, Kathleen Schmeler <KSchmele@mdanderson.org>,

Nadim Ajami <NAjami@mdanderson.org>,
"Wargo,Jennifer" <JWargo@mdanderson.org>, "Chapman,Bhavana
S" <BSChapman@mdanderson.org>, Jagannadha Sastry
<jsastry@mdanderson.org>, "Mezzari, Melissa" [redacted]
"Lin,Lilie L" <LLLin@mdanderson.org>, Lois Ramondetta
<lramonde@mdanderson.org>, Anuja Jhingran <ajhingra@mdanderson.org>,
Kathleen Schmeler <KSchmele@mdanderson.org>, Nadim Ajami
<NAjami@mdanderson.org>, "Wargo,Jennifer" <JWargo@mdanderson.org>,
[redacted]
Cc: Ann Klopp <AKlopp@mdanderson.org>, Lauren Colbert
<LColbert@mdanderson.org>, Ann Klopp <AKlopp@mdanderson.org>, Lauren
Colbert <LColbert@mdanderson.org>, "El Alam,Molly
B" <MBEl@mdanderson.org>
Subject: Manuscript - Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation

Hello all,
We have completed the first draft of the manuscript for our project “Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation”.

Mrs. El Alam, Dr. Colbert, Dr. Klopp and I have included you on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to *Nature Medicine*.

Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.

I am grateful for your time and feedback regarding this project! We hope to submit by 5/1/20.
Best,
Travis

Travis T. Sims, MD, MPH
Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

TITLE: Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation

Authors and Affiliations

Travis T. Sims, MD, MPH[1*], Molly B. El Alam, MPH[2*], Tatiana V. Karpinets, PhD[3], Kyoko Yoshida-Court, PhD[2], Xiaogang Wu, PhD[3], Greyson W. G. Biegert, BS[2], Andrea Y. Delgado Medrano, BS[2], Travis Solley, BS[2], Mustapha Ahmed-Kaddar, BS[2], Bhavana V. Chapman, MD[2], Jagannadha Sastry, PhD[4], Melissa P. Mezzari, PhD[5], Joseph F. Petrosino, PhD[5], Lilie L. Lin, MD[2], Lois Ramondetta, MD[1], Anuja Jhingran, MD[2], Kathleen M. Schmeler, MD[1], Nadim J Ajami, PhD[3], Jennifer Wargo, MD, MMSc[6], Lauren E. Colbert, MD, MSCR[2+], Ann H. Klopp, MD, PhD[2+]

[1]Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [2]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [3]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [4]Department of Immunology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [5]Department of Molecular Virology and Microbiology, Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA. [6]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

* Authors Contributed Equally

+Shared corresponding authorship

Correspondence: L.E. Colbert or A.H. Klopp, Department of Radiation Oncology, Unit

1422, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA. Telephone: 832-652-6033 (L.E.C.), 713-563-2444 (A.H.K); fax: 713-745-2398; e-mail: lcolbert@mdanderson.org, aklopp@mdanderson.org.

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Research Support

This research was supported in part by the National Institutes of Health (NIH) through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant 5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK).

Role of Funding Sources

The funding sources were not involved in the research hypothesis development, study design, data analysis, or manuscript writing. Data access was limited to the authors of this manuscript.

ABSTRACT

Background: Diversity of the gut microbiome is associated with response rates for patients with melanoma receiving immunotherapy and chemotherapy but has not been investigated in patients receiving radiation therapy. Additionally, studies investigating the gut microbiome and outcomes in cancer patients may not adjusted for established risk factors. We sought to determine if diversity and composition was independently associated with survival in cervical cancer (CC) patients receiving chemoradiation (CRT).

Methods: We analyzed baseline 16S rDNA fecal microbiomes of CC patients receiving standard CRT. Cervical tumor brushings were analyzed using flow cytometry. Patient and tumor characteristics were analyzed by univariate and multivariate Cox regression models for recurrence-free survival (RFS) and overall survival (OS) based on univariate p-value < 0.2. Characteristics included age, body mass index (BMI), race, stage, grade, histology, nodal status, and ~~max~~ tumor size. Alpha (within sample) diversity was evaluated using Shannon diversity index (SDI). Kaplan-Meier curves were generated for patients with high and normal BMI and overweight/obese BMI based on Cox analysis.

Results: 55 CC patients were included. Univariate analysis identified older age (Hazard Ratio (HR) of 0.93 (95% CI = 0.87-0.98, $P = 0.0096$)), SDI (HR of 0.51 (95% CI = 0.23-1.1, $P = 0.087$)) and BMI (HR of 0.92 (95% CI = 0.84-1, $P = 0.096$)) as risk factors for RFS. Multivariate survival analyses identified BMI and SDI as independent prognostic factors for RFS with a HR of 0.87 (95% CI = 0.77-0.98, $P = 0.02$) and 0.36 (95% CI = 0.15-0.84, $P = 0.018$) respectively. For OS,

multivariate survival analyses again identified BMI and SDI as independent prognostic factors with a HR of 0.78 (95% CI = 0.623-0.97, *P* = 0.025) and 0.19 (95% CI = 0.043-0.83, *P* = 0.028). For all patients, multiple taxa differed markedly between short term and long term survivors. Short term survivor fecal samples were significantly enriched in *porphyromonas, porphyromonadaceae, and dialister*, whereas long term survivor samples were significantly enriched in *Escherichia Shigella, Enterobacteriaceae, and Enterobacteriales* (P < 0.05; LDA score > 3.5) Analysis of cervical tumor brush flow cytometry revealed that patients with a high microbiome diversity had increased infiltration of CD4+ lymphocytes and well as activated subsets of CD4 cells expressing ki67+ and CD69+ over the course of radiation therapy.

Conclusion: Gut diversity is a significant predictor of OS in CC patients undergoing CRT and compositional differences were observed between patients who were short and long term survivors. Patients with high gut microbial diversity exhibit enhanced T cell signatures. Studies are needed to determine if modification of the gut microbiome will improve outcomes for women with cervical cancers.

Commented [MOU1]: Need to define

Key words: gynecologic cancer, microbiome, chemoradiation

INTRODUCTION

Cervical cancer continues to be one of the leading causes of cancer-associated mortality globally[1]. In the United States, more than 13,000 women ~~will be~~were diagnosed with invasive cervical cancer in 2019, resulting in more than 4,250 deaths[2]. Multimodality therapy consisting of concurrent chemoradiation (CRT) comprising external-beam radiotherapy (EBRT) and systemic chemotherapy followed by intracavitary brachytherapy continues to be the standard of care in clinical practice for locally advanced disease[3].

The fecal or gut microbiome, a diverse community of bacteria, archaea, fungi, protozoa, and viruses, is thought to influence host immunity by modulating multiple immunologic pathways, thus impacting health and disease[4–6]. Studies have suggested that dysbiosis of the gut microbiome confers a predisposition to certain malignancies and influences the body's response to a variety of cancer therapies, including chemotherapy, radiotherapy, and immunotherapy[6–10]. For example, melanoma patients are more likely to have a favorable response to immune checkpoint blockade and exhibit improved systemic and antitumor immunity if they have a more diverse intestinal microbiome[10].

Radiotherapy promotes the activation of T cells directed against tumor antigens[11–14]. In combination with immunotherapy, radiotherapy can maximize the antitumor immune response and promote durable disease control[15,16]. We theorize that the gut microbiota may modulate radioresponse through immunologic mechanisms[13,17]. Studies investigating the gut microbiome and outcomes in cancer patients often do not adjust for confounding patient and tumor characteristics. To assess this, we sought to identify independent gut microbial risk factors in cervical cancer (CC) patients receiving chemoradiation (CRT) and to evaluate their impact on

survival. We hypothesize that gut microbial differences may affect clinical outcomes in patients with cervical cancer.

RESULTS

Patient Characteristics

A total of 55 patients with a mean age of 47 years (range, 29-72 years) volunteered to participate in this study. The patients received standard treatment for cervical cancer with 5 weeks of EBRT and weekly cisplatin. After completion of EBRT, patients received brachytherapy. For evaluation of treatment response, patients underwent magnetic resonance imaging (MRI) at baseline and week 5 and positron emission tomography (PET)/computed tomography (CT) 3 months after treatment completion (Fig. 1a). Most patients had stage IIB disease (51%) and squamous histology (78%). Their clinicopathologic data are summarized in Supplementary Table 1. We staged cervical cancer using the 2014 International Federation of Gynecology and Obstetrics staging system. The median cervical tumor size according to MRI was 5.4 cm (range, 1.2-11.5 cm). Thirty patients (55%) had lymph node involvement according to PET or CT. We first analyzed the bacterial 16S rDNA (16Sv4) fecal microbiota at baseline with respect to disease histology, grade, and stage. We found that the baseline α-diversity (within tumor samples) and β-diversity (between samples) of the fecal microbiome in the cervical cancer patients did not differ according to histology, grade, or stage ($P > 0.05$) (Supplementary Fig. 1a-d).

Commented [MOU2]: MRI too? I would simplify and say imaging studies

Univariate and multivariate analysis of factors affecting recurrence free survival (RFS) and overall survival (OS)

Commented [MOU3]: Did you look at MDA vs LBJ?

In the univariate Cox proportional hazard regression model predicting RFS, 3 covariates showed $p \leq 0.2$. As shown in Table I, univariate analysis identified older age (Hazard Ratio (HR) of 0.93 (95% CI = 0.87-0.98, P = 0.0096)), SDI (HR of 0.51 (95% CI = 0.23-1.1, P = 0.087)) and BMI (HR of 0.92 (95% CI = 0.84-1, P = 0.096)) as risk factors for RFS. Multivariate survival analyses identified BMI and SDI as independent prognostic factors for RFS with a HR of 0.87 (95% CI = 0.77-0.98, P = 0.02) and 0.36 (95% CI = 0.15-0.84, P = 0.018) respectively. As shown in Table II2, univariate analysis identified SDI (HR of 0.34 (95% CI = 0.1-1.1, P = 0.08) and BMI (HR of 0.83 (95% CI = 0.69-1, P = 0.055)) as risk factors for OS. For OS, multivariate survival analyses again identified BMI and SDI as independent prognostic factors with a HR of 0.78 (95% CI = 0.623-0.97, P = 0.025) and 0.19 (95% CI = 0.043-0.83, P = 0.028) respectively.

Baseline Gut Microbiota Diversity is Associated with Favorable Responses

During the median follow-up period of 24.5 months, 7 patients died; all patients (12.7% of the total study population) died of disease (DOD). Figure 1 shows the Kaplan-Meier curves for RFS and OS. Given that in our univariate and multivariate analyses performed by Cox proportional hazard model Shannon index was confirmed as an independent predictor for RFS and OS, we first tested the relationship between diversity and RFS and OS in our cohort by stratifying patients based on high and low Shannon diversity metric. We stratified the patients by Shannon index as high-diversity versus low-diversity groups based on the cutoff value of Shannon index (2.69) calculated by receiver operating characteristic curve (ROC). We demonstrate that patients with high fecal alpha diversity at baseline showed a trend toward prolonged RFS and OS when compared to those with low diversity (P = 0.16 and 0.094, respectively) (Fig 1a,b). Next, because our univariate and multivariate analyses performed by Cox proportional hazard model also identified BMI as an independent predictor for RFS and OS we tested the relationship between

diversity and RFS and OS in our cohort by stratifying patients based on high and low Shannon diversity metric and normal or high BMI. As shown in Figure 1d,e, when BMI and gut diversity are stratified for at baseline, patients with normal BMI and higher SDI had a longer median RFS duration (P = 0.0027) (Fig 1d). OS (Fig 1e). Overall survival was longer for patients with normal BMI and higher gut diversity (P = 0.2).

Compositional Difference in Gut Microbiome in Response to chemoradiation

To further investigate whether the composition of gut microbiome was associated with the response to CRT, we used Linear discriminant analysis (LDA) Effect Size analysis to identify bacterial genera that were differentially enriched in short term and long term cervical cancer patients ($P < 0.05$; LDA score > 3.5). In all patients, multiple taxa differed significantly at baseline between short and long term survivors. Specifically, short term survivor fecal samples were significantly enriched in *porphyromonas*, *porphyromonadaceae, and dialister*, whereas long term survivor samples were significantly enriched in *Escherichia Shigella, Enterobacteriaceae, and Enterobacteriales* ($P < 0.05$; LDA score > 3.5, Fig 2a,b). Given that in our univariate analyses performed by Cox proportional hazard model *Pasteurellales, Haemophilus and Veillonella* were confirmed as an independent predictor for RFS and OS, we tested the relationship between these taxa and RFS and OS in our cohort by stratifying patients based on their relative abundance at baseline (Supplemental Fig 2). We demonstrate that patients with high relative abundance of *Veillonella* at baseline showed a trend toward prolonged RFS and OS when compared to those with a low relative abundance at baseline (P = 0.08 and P = 0.054, respectively).

Association between Gut Microbiota Profile and Immune Signatures

Because the gut microbiota is thought to influence disease progression partially through modulating systemic immune responses, we analyzed the cervical tumors in our cohort of patients via flow cytometry on tumor brushings performed before week 1, week 3 and week 5 of radiation therapy. To identify features associated with high gut diversity, Spearman correlation analysis was conducted between immune signatures at each time point. High Shannon diversity index was positively correlated with tumor infiltration of CD4 T cells at week 3, CD4ki67+ T-cells at week 5, (Table 3 and Fig 4a-d). The results suggest that patients with high gut diversity develop increased infiltration of activated CD4+ T-cell subsets.

Commented [MOU4]: Is Table 3 supplemental? Also – please be consistent in using roman numerals or not for the tables

DISCUSSION

The aim of this study was to identify independent gut microbial risk factors in cervical cancer patients receiving chemoradiation and to evaluate their impact on survival. We found higher BMI and gut diversity to be independent risk factors for RFS and OS in cervical cancer patients undergoing chemoradiation. The results indicate that overweight or obesity is a favorable prognostic factor independent of gut diversity. Additionally, our results demonstrate that patients with better clinical survival exhibit higher diversity as well as a distinct gut microbiome composition. Lastly the association between gut microbiome diversity and systematic immune signatures highlights helper CD4+ T cells as potential mediators of antitumor immunity upon CRT treatment.

Authors have previously described the gut microbiome and its effect on treatment outcomes for a variety of malignancies[39–41]. The diversity of gut microbiome is defined as the number and abundance distribution of distinct types of microorganisms colonizing within the gut[18]. In our study, higher alpha diversity at baseline correlated with an improved RFS and OS. High diversity implies more species harbor in the gut and suggests a difference in gut composition

between short term and long term survivors. Our results imply that the diversity of gut microbiota might be a shared benefit factor in those who respond well to CRT treatment. It is now generally accepted that the gut microbiome modulates immune responses, antitumor immunity, and clinical outcomes in a variety of malignancies[8,10,19]. The gut microbiome is thought to affect both innate and adaptive immune responses. Specifically how the gut microbiome exerts its influence continues to be explored, but this explanation may have important implications if specific taxa are found to change host response to treatment via immunomodulation[6]. In our study, T helper cell profiles at baseline correlate with gut diversity. These results confer that T cells and response to CRT are likely affected by the gut microbiota independent of other factors such as BMI. Using multi-color flow cytometry we performed correlation analysis on individual immune signatures and microbiota diversity. The frequency of helper CD4+ T cells were chiefly identified. Cervical cancer is considered to be an immunogenic tumor because its origin is dependent on a persistent infection with human papilloma-virus (HPV), most often HPV16 or HPV18[20]. Previous studies have reported that the number and functional orientation of tumor-infiltrating CD4+ and CD8+ T cells and the presence of M1 type macrophages strongly correlates with survival in patients with cervical cancer after chemoradiation[20,21]. T cells are capable of rapid antigen-specific responses and play critical roles in immune recall responses. In addition to the percentage of CD4+ t cell subsets, the increase in CD4 Ki67, CD4 CD69, and CD4 PD1 in the patients with high microbiota diversity implies that gut microbiome also modulates the proliferation of certain immune cell populations. Recent studies have already reported that chemoradiotherapy for cervical cancer induces unfavorable immune changes reflected by a decreased number of circulating lymphocytes, both CD4+ and CD8+ T cells, and an increased percentage in myeloid-cell populations, including myeloid-derived suppressor cells and monocytes[20]. Whereas CD4+ T cells infiltrating in tumor

220 microenvironment are thought to help the activity of other immune cells by releasing T cell
221 cytokines, circulating CD4+ T cell subsets reported here are probably inclined to reflect the role
222 of gut microbiota on systemic immune responses. How peripheral memory CD4+ T cell signatures
223 affect the efficacy of CRT treatment needs to be investigated in the future. Our study shows that
224 the diversity of gut microbiota is associated with favorable response to CRT against cervical
225 cancer. Considering the correlation between microbiota diversity and peripheral helper T cells
226 being reshaped upon CRT treatment, we propose that patients with more diverse gut microbiota at
227 baseline may benefit from CRT to a greater extent. This might be mediated by reprogramming
228 systemic antitumor immune responses. The significance of our study lies in that the modulation of
229 gut microbiota before treatment might provide an alternative way to enhance the efficacy of CRT,
230 specifically in cases with positive lymph nodes and advanced stages in which systemic failure of
231 current therapies represents a major challenge. Our results suggest that changes in the gut
232 microenvironment contribute substantially to treatment success or failure, particularly in so-called
233 immunogenic tumors like cervical cancer. Additionally, there is emerging data describing the
234 influence of the gut microbiome as it pertains to radiotherapy[22]. Given that radiation can change
235 the composition of the gut microbiome by altering the relative abundance of different taxa, we
236 have to postulate whether it is these changes that ultimately alter the effectiveness of radiotherapy
237 for cervical cancer[6,23,24].

238 In our cohort, at baseline, a higher relative abundance of *Veillonella* resulted in a trend
239 toward prolonged RFS and OS. Our own group has previously characterized the 16S rDNA fecal
240 microbiome cervical cancer patients compared to healthy female controls, and have reported on
241 differences in the relative abundance of specific taxa[42]. Our new findings support the hypothesis
242 that organisms like *Veillonella* inhabiting the gut microbiome may be manipulated to improve

cancer treatment response. Knowing specific gut microbial organisms that inhabit and undergo changes in patients with cervical cancer during CRT provides further insight into mechanisms that may modulate immune response and potentiate treatment outcomes in cancer patients. The results of our study illustrate the potential of intentionally modifying the gut microbiota to accumulate CRT-tolerant species as an interventional strategy to enhance response of cervical cancer to CRT. Researchers have studied the treatment-enhancing utility of the gut microbiota in multiple areas of medicine[9,38]. For example, human fecal microbial transplants have protected germ-free mice from arsenic-induced mortality and reduced the number of antibiotic-resistant genes in patients with recurrent *Clostridium difficile* infections[39,40]. Also, Wang *et al.*[41] recently reported on the first case series of patients with immune checkpoint inhibitor-associated colitis successfully treated with fecal microbiota transplantation. With respect to how the gut microbiome can modulate the host response to chemotherapy, a previous review highlighted three important clinical elements: facilitation of drug efficacy, compromise of anticancer effects, and mediation of toxicity[42]. The authors went on to predict how the gut microbiome could be modified in clinical practice to increase cancer treatment efficacy and reduce toxicity. For example, in a murine model, radiation-induced dysbiosis increased the susceptibility of mice to radiotherapy-related gastrointestinal toxic effects[23]. Determining whether changes in the human gut microbiome during CRT affect patients' risk of treatment-related toxic effects may be an area for further investigation.

The "obesity paradox", which suggest a positive effects of increasing BMI as it pertains to a specific disease, was firstly reported in heart failure[25], but has since been described in a variety of disease processes including coronary artery disease, kidney disease, diabetes, and a variety of malignancies, including other gynecologic cancers[26–28]. Theories centered around the "obesity paradox" suggest that patients with a high BMI may be better able to withstand cancer-induced

consumption and stress compared with patients with a low BMI[37]. Other theories include greater metabolic reserve, an attenuated response to hormones involved in the renin–angiotensin–aldosterone system, fitness and its association with adiposity and clinical prognosis, and unmeasured confounding factors[26]. For example, in uterine cancer it has been reported that the risk of recurrence differed significantly by BMI[29]. Specifically, a greater proportion of obese women (BMI ≥ 40) met criteria for having a low risk of recurrence, while thin women tended to have a high intermediate risk or recurrence. There have been many studies investigating the impact of BMI on cervical cancer, but the association between BMI and cervical cancer remains unclear[30]. Most cervical cancer is caused by a persistent infection with a high risk ~~human papillomavirus (~~HPV~~)~~. However, it has been suggested that obesity may increase the risk of cervical cancer[31]. Contributing factors include poor screening and that body fat distribution hormonally influences the risk of glandular cervical carcinoma like adenocarcinoma of the cervix[32,33].

In contrast, Brinton *et al.* reported that body weight was not an independent prognostic factor for squamous cell tumors, and a slight increased risk of adenocarcinoma, although this was not significant[34]. Tornberg *et al.* reported that there was not a significant relationship between being overweight and cervical cancer[35] and a review conducted in 2008 by Lane *et al.* did not report a relationship between cervical cancer and obesity siting a of a lack of evidence[36]. Finally, a meta-analysis done by Poorolajal *et al.* in 2016 indicated that being overweight (BMI 25–29.9 kg/m2), is not associated with an increased risk of cervical cancer, but that obesity (BMI ≥30 kg/m2) is weakly associated with an increased risk of cervical cancer[30]. However, the authors warned that more evidence, based on large prospective cohort studies, is required to provide conclusive evidence on whether or not BMI is associated with an increased risk of cervical cancer. These factors demonstrate the need to better understand if and how obesity increases cervical

Commented [MOU5]: Define these categories earlier and then don't need to repeat

cancer risk. The inconsistent conclusions among studies investigating the association between BMI and cervical cancer may be attributed to numerous factors including, but not limited to, patient selection criteria, sample size and generalizability of the study population to the general public. Among these factors, patient selection criteria may be especially important, because tumor histology seems to be closely associated with BMI.

The strengths of this study include the use of careful clinical staging, histopathology, and reliable phylogenetic and statistical analysis to assess bacterial community compositional changes using both microbial divergence and taxon-based methods. Additionally, we followed a complete protocol for 16S sequencing ranging from the sample collection method to DNA extraction and sequencing, thus limiting artifactual variations. Although this study yielded intriguing findings, it was limited by its small sample. Consequently, the sample size limited our ability to weigh statistical power. However, results presented herein provide solid evidence of the effect of CRT on the gut microbiome.

In conclusion, our study demonstrated that gut diversity is a significant factor for predicting OS in CC patients undergoing CRT when BMI is accounted for, and may help explain the "obesity paradox" in cancer response. Our study shows that the diversity of gut microbiota is associated with a favorable response to chemoradiation against cervical cancer. Considering the correlation between microbiota diversity and T cells being influenced with CRT treatment, patients with more diverse gut microbiota at baseline may benefit from CRT to a greater extent. The significance of our study lies in that the modulation of gut microbiota before CRT might provide an alternative way to enhance the efficacy of CRT but this needs to be validated in large cohort studies. Studies exploring the relationship between gut diversity, CRT, and treatment efficacy are needed to further understand the role of the gut microbiome in treatment outcomes.

ONLINE METHODS

Participants and clinical data. Gut microbiome and cervical swab samples were collected prospectively from cervical cancer patients according to a protocol approved by The University of Texas MD Anderson Cancer Center Institutional Review Board (MDACC 2014-0543) for patients with biopsy-proven carcinoma of the cervix treated at MD Anderson and the Lyndon B. Johnson Hospital Oncology Clinic from September 22, 2015, to January 11, 2019. All patients had new diagnoses of locally advanced, nonmetastatic carcinoma of the cervix and underwent definitive CRT with EBRT followed by brachytherapy. Patients received a minimum of 45 Gy via EBRT in 25 fractions over 5 weeks with weekly cisplatin followed by two brachytherapy sessions at approximately weeks 5 and 7 with EBRT in between for gross nodal disease or persistent disease in the parametrium. Patients with stage IB1 cancer were given CRT due to the presence of nodal disease. Clinical variables, demographics, and pathologic reports were abstracted from electronic medical records.

Sample collection and DNA extraction. Stool was collected from all patients by a clinician performing rectal exams at five time points (baseline; weeks 1, 3, and 5 of radiotherapy; and 3 months after CRT completion) using a matrix-designed quick-release Isohelix swab to characterize the diversity and composition of the microbiome over time. The swabs were stored in 20 µl of protease K and 400 µl of lysis buffer (Isohelix) and kept at -80°C within 1 h of sample collection.

16S rRNA gene sequencing and sequence data processing. 16S rRNA sequencing was performed for fecal samples obtained from all patients at four time points to characterize the

diversity and composition of the microbiome over time. 16S rRNA gene sequencing was done at the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine. 16S rRNA was sequenced using approaches adapted from those used for the Human Microbiome Project[43]. The 16S rDNA V4 region was amplified via polymerase chain reaction with primers that contained sequencing adapters and single-end barcodes, allowing for pooling and direct sequencing of polymerase chain reaction products. Amplicons were sequenced on the MiSeq platform (Illumina) using the 2 x 250-bp paired-end protocol, yielding paired-end reads that overlapped nearly completely. Sequence reads were demultiplexed, quality-filtered, and subsequently merged using the USEARCH sequence analysis tool (version 7.0.1090) (4). 16S rRNA gene sequences were bundled into operational taxonomic units at a similarity cutoff value of 97% using the UPARSE algorithm[44]. To generate taxonomies, operational taxonomic units were mapped to an enhanced version of the SILVA rRNA database containing the 16Sv4 region. A custom script was used to create an operational taxonomic unit table from the output files generated as described above for downstream analyses of α-diversity, β-diversity, and phylogenetic trends. Principal coordinates analysis was performed by institution and sample set to make certain no batch effects were present.

Flow Cytometry. Immunostaining was performed according to standard protocols. Cells were fixed using the Foxp3/Transcription Factor Staining Buffer Set (eBioscience) and stained with a 16 color panel with antibodies from Biolegend, BD Bioscience, eBioscience, and Life Technologies. Analysis was performed on a 5-laser, 18 color LSRFortessa X-20 Flow Cytometer (BD Biosciences). Analysis was performed on Flowjo Software (INFO). Briefly, cells were stained with intracellular mAB for 30 minutes at 4C in the presence of anti-Cd16/Cd32 mAB (BD

Bioscience), fixed with Foxp3/Transcription Factor Staining Buffer Set (eBioscience), and held in FACS (Corning, 2 mM EDTA, 2% FBS). Counting beads (Thermo Fisher) were used for single color controls.

Statistical analyses. For microbiome analysis, rarefaction depth was set at 7066 reads. The ISD index was used to evaluate α-diversity (within samples), and principle coordinates analysis of unweighted UniFrac distances was used to examine β-diversity (between samples). Patient and tumor characteristics were analyzed by univariate and multivariate Cox regression models for Recurrence-free survival (RFS) and Overall survival (OS) based on univariate p-value < 0.2. Characteristics included age, body mass index (BMI), race, stage, grade, histology, nodal status, smoking status, antibiotic use and max tumor size. For each outcome of interest, a multivariate Cox regression analysis was performed to adjust for the effects of prognostic factors identified on univariate analysis as influencing survival in cervical cancer. These analyses were conducted using covariates with $p \leq 0.2$ in a stepwise fashion. Alpha (within sample) diversity was evaluated using Shannon diversity index (SDI). The relative abundance of microbial taxa, classes, and genera was determined using LDA Effect Size[45], applying the one-against-all strategy with a threshold of 2 for the logarithmic LDA score for discriminative features and α of 0.05 for factorial Kruskal-Wallis testing among classes. LDA Effect Size analysis was restricted to bacteria present in 20% or more of the study population. Kaplan-Meier curves were generated for patients with normal BMI and overweight/obese BMI based on Cox analysis and clostridia abundance. The significance of differences was determined using the log-rank test. Gut microbial diversity, RFS, and OS were also compared using Pearson correlation, linear regression, and Cox regression analysis.

ACKNOWLEDGEMENTS

This work was supported in part by the Radiological Society of North America Resident/Fellow Award (to L.E.C.), the NIH/NCI under award number P30CA016672, and an NIH T32 grant (#5T32 CA101642-14; to T.T.S.). This study was partially funded by the MD Anderson HPV-Related Cancers Moon Shot program (to L.E.C. and A.K.). The human subjects who participated in this study are gratefully acknowledged.

AUTHOR CONTRIBUTION

All authors were involved in subject identification, data collection, interpretation of the statistical analysis, and review and approval of the final manuscript. The study concept was conceived by L.E.C., A.K., and T.T.S. The manuscript was written by T.T.S.

COMPETING INTERESTS

The authors report no conflicts of interest, financial or otherwise, related to the subject of this article.

ROLE OF FUNDING SOURCES

The funding sources were not involved in the development of the research hypothesis, study design, data analysis, or writing of the manuscript. Data access was limited to the authors of the manuscript.

References:

1. Islami, F., Fedewa, S. A. & Jemal, A. Trends in cervical cancer incidence rates by age, race/ethnicity, histological subtype, and stage at diagnosis in the United States. *Prev. Med.* **123**, 316–323 (2019).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* **69**, 7–34 (2019).
3. Monk, B. J., Tewari, K. S. & Koh, W.-J. Multimodality therapy for locally advanced cervical carcinoma: state of the art and future directions. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25**, 2952–2965 (2007).
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
5. Cho, J. A. & Chinnapen, D. J. F. Targeting friend and foe: Emerging therapeutics in the age of gut microbiome and disease. *J. Microbiol.* **56**, 183–188 (2018).
6. Helmink, B. A., Khan, M. A. W., Hermann, A., Gopalakrishnan, V. & Wargo, J. A. The microbiome, cancer, and cancer therapy. *Nat. Med.* **25**, 377–388 (2019).
7. Cerf-Bensussan, N. & Gaboriau-Routhiau, V. The immune system and the gut microbiota: friends or foes? *Nat. Rev. Immunol.* **10**, 735–744 (2010).
8. McQuade, J. L., Daniel, C. R., Helmink, B. A. & Wargo, J. A. Modulating the microbiome to improve therapeutic response in cancer. *Lancet Oncol.* **20**, e77–e91 (2019).
9. Roy, S. & Trinchieri, G. Microbiota: a key orchestrator of cancer therapy. *Nat. Rev. Cancer* **17**, 271–285 (2017).
10. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018).
11. Dewan, M. Z. *et al.* Fractionated but Not Single-Dose Radiotherapy Induces an Immune-Mediated Abscopal Effect when Combined with Anti-CTLA-4 Antibody. *Clin. Cancer Res.* **15**, 5379–5388 (2009).
12. Lugade, A. A. *et al.* Local Radiation Therapy of B16 Melanoma Tumors Increases the Generation of Tumor Antigen-Specific Effector Cells That Traffic to the Tumor. *J. Immunol.* **174**, 7516–7523 (2005).
13. Lhuillier, C., Rudqvist, N.-P., Elemento, O., Formenti, S. C. & Demaria, S. Radiation therapy and anti-tumor immunity: exposing immunogenic mutations to the immune system. *Genome Med.* **11**, 40 (2019).
14. Lee, Y. *et al.* Therapeutic effects of ablative radiation on local tumor require CD8+ T cells: changing strategies for cancer treatment. *Blood* **114**, 589–595 (2009).
15. Ngwa, W. *et al.* Using immunotherapy to boost the abscopal effect. *Nat. Rev. Cancer* **18**, 313–322 (2018).
16. Grassberger, C., Ellsworth, S. G., Wilks, M. Q., Keane, F. K. & Loeffler, J. S. Assessing the interactions between radiotherapy and antitumour immunity. *Nat. Rev. Clin. Oncol.* **16**, 729–745 (2019).
17. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).
18. Jin, Y. *et al.* The Diversity of Gut Microbiome is Associated With Favorable Responses to Anti-Programmed Death 1 Immunotherapy in Chinese Patients With NSCLC. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **14**, 1378–1389 (2019).
19. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).

20. Meir, H. van *et al.* Impact of (chemo)radiotherapy on immune cell composition and function in cervical cancer patients. *OncoImmunology* **6**, e1267095 (2017).
21. Jordanova, E. S. *et al.* Human leukocyte antigen class I, MHC class I chain-related molecule A, and CD8+/regulatory T-cell ratio: which variable determines survival of cervical cancer patients? *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **14**, 2028–2035 (2008).
22. Muls, A. *et al.* Systematic Review: The Impact of Cancer Treatment on the Gut and Vaginal Microbiome in Women With a Gynecological Malignancy. *Int. J. Gynecol. Cancer* **27**, 1550–1559 (2017).
23. Gerassy-Vainberg, S. *et al.* Radiation induces proinflammatory dysbiosis: transmission of inflammatory susceptibility by host cytokine induction. *Gut* **67**, 97–107 (2018).
24. Sokol, H. & Adolph, T. E. The microbiota: an underestimated actor in radiation-induced lesions? *Gut* **67**, 1–2 (2018).
25. Lavie, C. J., Osman, A. F., Milani, R. V. & Mehra, M. R. Body composition and prognosis in chronic systolic heart failure: the obesity paradox. *Am. J. Cardiol.* **91**, 891–894 (2003).
26. Lavie, C. J., De Schutter, A. & Milani, R. V. Healthy obese versus unhealthy lean: the obesity paradox. *Nat. Rev. Endocrinol.* **11**, 55–62 (2015).
27. Walker, J. L. *et al.* Laparoscopy Compared With Laparotomy for Comprehensive Surgical Staging of Uterine Cancer: Gynecologic Oncology Group Study LAP2. *J. Clin. Oncol.* **27**, 5331–5336 (2009).
28. S, A.-A., A, S. & J, K. Obesity and cancer. *J. Med.* **33**, 73–86 (2002).
29. Gunderson, C. C., Java, J., Moore, K. N. & Walker, J. L. The impact of obesity on surgical staging, complications, and survival with uterine cancer: A Gynecologic Oncology Group LAP2 ancillary data study. *Gynecol. Oncol.* **133**, 23–27 (2014).
30. Poorolajal, J. & Jenabi, E. The association between BMI and cervical cancer risk: a meta-analysis. *Eur. J. Cancer Prev. Off. J. Eur. Cancer Prev. Organ. ECP* **25**, 232–238 (2016).
31. Nagaraju, G. P. *et al.* The role of adiponectin in obesity-associated female-specific carcinogenesis. *Cytokine Growth Factor Rev.* **31**, 37–48 (2016).
32. Martín-López, R. *et al.* Breast and cervical cancer screening in Spain and predictors of adherence. *Eur. J. Cancer Prev.* **19**, 239–245 (2010).
33. Lacey, J. V. *et al.* Obesity as a potential risk factor for adenocarcinomas and squamous cell carcinomas of the uterine cervix. *Cancer* **98**, 814–821 (2003).
34. Brinton, L. A. *et al.* Risk Factors for Cervical Cancer by Histology. *Gynecol. Oncol.* **51**, 301–306 (1993).
35. Törnberg, S. A. & Carstensen, J. M. Relationship between Quetelet's index and cancer of breast and female genital tract in 47,000 women followed for 25 years. *Br. J. Cancer* **69**, 358–361 (1994).
36. Lane, G. Obesity and gynaecological cancer. *Menopause Int.* **14**, 33–37 (2008).
37. Chen, H.-N. *et al.* The Impact of Body Mass Index on the Surgical Outcomes of Patients With Gastric Cancer: A 10-Year, Single-Institution Cohort Study. *Medicine (Baltimore)* **94**, e1769 (2015).
38. Kohl, K. D., Weiss, R. B., Cox, J., Dale, C. & Denise Dearing, M. Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecol. Lett.* **17**, 1238–1246 (2014).
39. Coryell, M., McAlpine, M., Pinkham, N. V., McDermott, T. R. & Walk, S. T. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. *Nat. Commun.* **9**, 5424 (2018).

40. Millan, B. *et al.* Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent *Clostridium difficile* Infection. *Clin. Infect. Dis.* **62**, 1479–1486 (2016).
41. Wang, Y. *et al.* Fecal microbiota transplantation for refractory immune checkpoint inhibitor-associated colitis. *Nat. Med.* **24**, 1804–1808 (2018).
42. Alexander, J. L. *et al.* Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 356–365 (2017).
43. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
44. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
45. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

508

Date : 4/21/2020 11:33:14 AM
From : "Sastry,Jagannadha K" jsastry@mdanderson.org
To : "Sims,Travis T." TTSims@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Schmeler,Kathleen M" KSchmele@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "Chapman,Bhavana S" BSChapman@mdanderson.org, "Mezzari, Melissa" [redacted], "Lin,Lilie L" LLLin@mdanderson.org, "Ramondetta,Lois M" lramonde@mdanderson.org, "Jhingran,Anuja" ajhingra@mdanderson.org, "Schmeler,Kathleen M" KSchmele@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "[redacted]
Cc : "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Klopp,Ann H" AKlopp@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org
Subject : RE: Manuscript - Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation
Attachment : Manuscript - Gut microbiome diversity as an independent predictor of survival in cervical cancer patients receiving chemoradiation V1 - JS edits.docx;

Please see my edits. At many places in the results and discussion sections, the flow analyses performed for immune responses is once again referred to as “systemic response” as opposed to “intra-tumoral” or “local”. I corrected this at a couple of places and highlighted in yellow at other places. Please pay attention to streamline this properly to remove confusion. Feel free to let me know if you need further assistance

Best
Jagan Sastry

_____

Jagannadha Sastry, Ph. D.
Member, UT Health, MD Anderson Cancer Center- Graduate School of Biomedical Sciences
Professor, Department of Thoracic Head & Neck Medical Oncology and Department of Comparative Medicine
The University of Texas MD Anderson Cancer Center
Room: T8.3908
1515 Holcomb Blvd, Houston, TX 77030
Phone: 713 563 3304; Cell: 281 787 5689
E-mail: jsastry@mdanderson.org

From: Sims,Travis T.
Sent: Wednesday, April 15, 2020 7:21 PM
To: Sastry,Jagannadha K <jsastry@mdanderson.org>; Karpinets,Tatiana V

<TVKarpinets@mdanderson.org>; Lin,Lilie L <LLLin@mdanderson.org>; Ramondetta,Lois M <lramonde@mdanderson.org>; Jhingran,Anuja <ajhingra@mdanderson.org>; Schmeler,Kathleen M <KSchmele@mdanderson.org>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>; Chapman,Bhavana S <BSChapman@mdanderson.org>; Sastry,Jagannadha K <jsastry@mdanderson.org>; Mezzari, Melissa <[redacted]>; Lin,Lilie L <LLLin@mdanderson.org>; Ramondetta,Lois M <lramonde@mdanderson.org>; Jhingran,Anuja <ajhingra@mdanderson.org>; Schmeler,Kathleen M <KSchmele@mdanderson.org>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>;
[redacted]
Cc: Klopp,Ann H <AKlopp@mdanderson.org>; Colbert,Lauren Elizabeth <LColbert@mdanderson.org>; Klopp,Ann H <AKlopp@mdanderson.org>; Colbert,Lauren Elizabeth <LColbert@mdanderson.org>; El Alam,Molly B <MBEl@mdanderson.org>
Subject: Manuscript - Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation

Hello all,

We have completed the first draft of the manuscript for our project “Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation”.

Mrs. El Alam, Dr. Colbert, Dr. Klopp and I have included you on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to *Nature Medicine*.

Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.

I am grateful for your time and feedback regarding this project! We hope to submit by 5/1/20.

Best,
Travis

Travis T. Sims, MD, MPH
Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

TITLE: Gut microbiome diversity is an independent predictor of survival in cervical cancer patients receiving chemoradiation

Authors and Affiliations

Travis T. Sims, MD, MPH[1*], Molly B. El Alam, MPH[2*], Tatiana V. Karpinets, PhD[3], Kyoko Yoshida-Court, PhD[2], Xiaogang Wu, PhD[3], Greyson W. G. Biegert, BS[2], Andrea Y. Delgado Medrano, BS[2], Travis Solley, BS[2], Mustapha Ahmed-Kaddar, BS[2], Bhavana V. Chapman, MD[2], K. Jagannadha Sastry, PhD[4], Melissa P. Mezzari, PhD[5], Joseph F. Petrosino, PhD[5], Lilie L. Lin, MD[2], Lois Ramondetta, MD[1], Anuja Jhingran, MD[2], Kathleen Schmeler, MD[1], Nadim J Ajami, PhD[3], Jennifer Wargo, MD, MMSc[6], Lauren E. Colbert, MD, MSCR[2+], Ann H. Klopp, MD, PhD[2+]

[1]Department of Gynecologic Oncology and Reproductive Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [2]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [3]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [4]Department of ~~Immunology~~Thoracic Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [5]Department of Molecular Virology and Microbiology, Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA. [6]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

* Authors Contributed Equally

+Shared corresponding authorship

Commented [SK1]: Include the people below from my group who performed the multi-color flow cytometry

Stephanie Dorta-Estremera, Venkatesh L. Hegde, and Sita Nookala

Commented [SK2]: Please also include my affiliation with the graduate school as below:
The University of Texas MD Anderson Cancer Center
UTHealth Graduate School of Biomedical Sciences at Houston, Houston, TX 77030

Correspondence: L.E. Colbert or A.H. Klopp, Department of Radiation Oncology, Unit 1422, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA. Telephone: 832-652-6033 (L.E.C.), 713-563-2444 (A.H.K); fax: 713-745-2398; e-mail: lcolbert@mdanderson.org, aklopp@mdanderson.org.

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Research Support

This research was supported in part by the National Institutes of Health (NIH) through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant 5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK).

Role of Funding Sources

The funding sources were not involved in the research hypothesis development, study design, data analysis, or manuscript writing. Data access was limited to the authors of this manuscript.

ABSTRACT

Background: Diversity of the gut microbiome is associated with response rates for patients with melanoma receiving immunotherapy and chemotherapy but has not been investigated in patients receiving radiation therapy. Additionally, studies investigating the gut microbiome and outcomes in cancer patients may not adjusted for established risk factors. We sought to determine if diversity and composition was independently associated with survival in cervical cancer (CC) patients receiving chemoradiation (CRT).

Methods: We analyzed baseline 16S rDNA fecal microbiomes of CC patients receiving standard CRT. Immune cells isolated from the c~~C~~ervical tumor brushings were analyzed using flow cytometry. Patient and tumor characteristics were analyzed by univariate and multivariate Cox regression models for recurrence-free survival (RFS) and overall survival (OS) based on univariate p-value < 0.2. Characteristics included age, body mass index (BMI), race, stage, grade, histology, nodal status, and max tumor size. Alpha (within sample) diversity was evaluated using Shannon diversity index (SDI). Kaplan-Meier curves were generated for patients with high and normal BMI and overweight/obese BMI based on Cox analysis.

Results: 55 CC patients were included. Univariate analysis identified older age (Hazard Ratio (HR) of 0.93 (95% CI = 0.87-0.98, $P = 0.0096$)), SDI (HR of 0.51 (95% CI = 0.23-1.1, $P = 0.087$)) and BMI (HR of 0.92 (95% CI = 0.84-1, $P = 0.096$)) as risk factors for RFS. Multivariate survival analyses identified BMI and SDI as independent prognostic factors for RFS with a HR of 0.87

(95% CI = 0.77-0.98, *P* = 0.02) and 0.36 (95% CI = 0.15-0.84, *P* = 0.018) respectively. For OS, multivariate survival analyses again identified BMI and SDI as independent prognostic factors with a HR of 0.78 (95% CI = 0.623-0.97, *P* = 0.025) and 0.19 (95% CI = 0.043-0.83, *P* = 0.028) For all patients, multiple taxa differed markedly between short term and long term survivors. Short term survivor fecal samples were significantly enriched in *porphyromonas, porphyromonadaceae, and dialister*, whereas long term survivor samples were significantly enriched in *Escherichia Shigella, Enterobacteriaceae, and Enterobacteriales* (P < 0.05; LDA score > 3.5) Analysis of immune cells from cervical tumor brush samples by flow cytometry revealed that patients with a high microbiome diversity had increased infiltration of CD4+ lymphocytes as~~nd~~ well as activated subsets of CD4 cells expressing ki67+ and CD69+ over the course of radiation therapy.

Conclusion: Gut diversity is a significant predictor of OS in CC patients undergoing CRT and compositional differences were observed between patients who were short and long term survivors. Patients with high gut microbial diversity exhibit enhanced T cell signatures. Studies are needed to determine if modification of the gut microbiome will improve outcomes for women with cervical cancers.

Key words: gynecologic cancer, microbiome, chemoradiation

INTRODUCTION

Cervical cancer continues to be one of the leading causes of cancer-associated mortality globally[1]. In the United States, more than 13,000 women will be diagnosed with invasive cervical cancer in 2019, resulting in more than 4,250 deaths[2]. Multimodality therapy consisting of concurrent chemoradiation (CRT) comprising external-beam radiotherapy (EBRT) and systemic chemotherapy followed by intracavitary brachytherapy continues to be the standard of care in clinical practice for locally advanced disease[3].

The fecal or gut microbiome, a diverse community of bacteria, archaea, fungi, protozoa, and viruses, is thought to influence host immunity by modulating multiple immunologic pathways, thus impacting health and disease[4–6]. Studies have suggested that dysbiosis of the gut microbiome confers a predisposition to certain malignancies and influences the body's response to a variety of cancer therapies, including chemotherapy, radiotherapy, and immunotherapy[6–10]. For example, melanoma patients are more likely to have a favorable response to immune checkpoint blockade and exhibit improved systemic and antitumor immunity if they have a more diverse intestinal microbiome[10].

Radiotherapy promotes the activation of T cells directed against tumor antigens[11–14]. In combination with immunotherapy, radiotherapy can maximize the antitumor immune response and promote durable disease control[15,16]. We theorize that the gut microbiota may modulate radioresponse through immunologic mechanisms[13,17]. Studies investigating the gut microbiome and outcomes in cancer patients often do not adjust for confounding patient and tumor characteristics. To assess this, we sought to identify independent gut microbial risk factors in cervical cancer (CC) patients receiving chemoradiation (CRT) and to evaluate their impact on

survival. We hypothesize that gut microbial differences may affect clinical outcomes in patients with cervical cancer.

RESULTS

Patient Characteristics

A total of 55 patients with a mean age of 47 years (range, 29-72 years) volunteered to participate in this study. The patients received standard treatment for cervical cancer with 5 weeks of EBRT and weekly cisplatin. After completion of EBRT, patients received brachytherapy. For evaluation of treatment response, patients underwent magnetic resonance imaging (MRI) at baseline and week 5 and positron emission tomography (PET)/computed tomography (CT) 3 months after treatment completion (Fig. 1a). Most patients had stage IIB disease (51%) and squamous histology (78%). Their clinicopathologic data are summarized in Supplementary Table 1. We staged cervical cancer using the 2014 International Federation of Gynecology and Obstetrics staging system. The median cervical tumor size according to MRI was 5.4 cm (range, 1.2-11.5 cm). Thirty patients (55%) had lymph node involvement according to PET or CT. We first analyzed the bacterial 16S rDNA (16Sv4) fecal microbiota at baseline with respect to disease histology, grade, and stage. We found that the baseline α-diversity (within tumor samples) and β-diversity (between samples) of the fecal microbiome in the cervical cancer patients did not differ according to histology, grade, or stage ($P > 0.05$) (Supplementary Fig. 1a-d).

Univariate and multivariate analysis of factors affecting recurrence free survival (RFS) and overall survival (OS)

In the univariate Cox proportional hazard regression model predicting RFS, 3 covariates showed $p \leq 0.2$. As shown in Table I, univariate analysis identified older age (Hazard Ratio (HR) of 0.93 (95% CI = 0.87-0.98, P = 0.0096)), SDI (HR of 0.51 (95% CI = 0.23-1.1, P = 0.087)) and BMI (HR of 0.92 (95% CI = 0.84-1, P = 0.096)) as risk factors for RFS. Multivariate survival analyses identified BMI and SDI as independent prognostic factors for RFS with a HR of 0.87 (95% CI = 0.77-0.98, P = 0.02) and 0.36 (95% CI = 0.15-0.84, P = 0.018) respectively. As shown in Table 2, univariate analysis identified SDI (HR of 0.34 (95% CI = 0.1-1.1, P = 0.08) and BMI (HR of 0.83 (95% CI = 0.69-1, P = 0.055)) as risk factors for OS. For OS, multivariate survival analyses again identified BMI and SDI as independent prognostic factors with a HR of 0.78 (95% CI = 0.623-0.97, P = 0.025) and 0.19 (95% CI = 0.043-0.83, P = 0.028) respectively.

Baseline Gut Microbiota Diversity is Associated with Favorable Responses

During the median follow-up period of 24.5 months, 7 patients died; all patients (12.7% of the total study population) died of disease (DOD). Figure 1 shows the Kaplan-Meier curves for RFS and OS. Given that in our univariate and multivariate analyses performed by Cox proportional hazard model Shannon index was confirmed as an independent predictor for RFS and OS, we first tested the relationship between diversity and RFS and OS in our cohort by stratifying patients based on high and low Shannon diversity metric. We stratified the patients by Shannon index as high-diversity versus low-diversity groups based on the cutoff value of Shannon index (2.69) calculated by receiver operating characteristic curve (ROC). We demonstrate that patients with high fecal alpha diversity at baseline showed a trend toward prolonged RFS and OS when compared to those with low diversity (P = 0.16 and 0.094, respectively) (Fig 1a,b). Next, because our univariate and multivariate analyses performed by Cox proportional hazard model also identified BMI as an independent predictor for RFS and OS we tested the relationship between

diversity and RFS and OS in our cohort by stratifying patients based on high and low Shannon diversity metric and normal or high BMI. As shown in Figure 1d,e, when BMI and gut diversity are stratified for at baseline, patients with normal BMI and higher SDI had a longer median RFS duration (P = 0.0027) (Fig 1d). OS (Fig 1e). Overall survival was longer for patients with normal BMI and higher gut diversity (P = 0.2).

Compositional Difference in Gut Microbiome in Response to chemoradiation

To further investigate whether the composition of gut microbiome was associated with the response to CRT, we used Linear discriminant analysis (LDA) Effect Size analysis to identify bacterial genera that were differentially enriched in short term and long term cervical cancer patients ($P < 0.05$; LDA score > 3.5). In all patients, multiple taxa differed significantly at baseline between short and long term survivors. Specifically, short term survivor fecal samples were significantly enriched in *porphyromonas*, *porphyromonadaceae, and dialister*, whereas long term survivor samples were significantly enriched in *Escherichia Shigella, Enterobacteriaceae, and Enterobacteriales* ($P < 0.05$; LDA score > 3.5, Fig 2a,b). Given that in our univariate analyses performed by Cox proportional hazard model *Pasteurellales, Haemophilus and Veillonella* were confirmed as an independent predictor for RFS and OS, we tested the relationship between these taxa and RFS and OS in our cohort by stratifying patients based on their relative abundance at baseline (Supplemental Fig 2). We demonstrate that patients with high relative abundance of *Veillonella* at baseline showed a trend toward prolonged RFS and OS when compared to those with a low relative abundance at baseline (P = 0.08 and P = 0.054, respectively).

Association between Gut Microbiota Profile and Immune Signatures

Because the gut microbiota is thought to influence disease progression partially through modulating systemic immune responses, we analyzed the cervical tumors in our cohort of patients via flow cytometry on tumor brushings performed before week 1, week 3 and week 5 of radiation therapy. To identify features associated with high gut diversity, Spearman correlation analysis was conducted between immune signatures at each time point. High Shannon diversity index was positively correlated with tumor infiltration of CD4 T cells at week 3, CD4ki67+ T-cells at week 5, (Table 3 and Fig 4a-d). The results suggest that patients with high gut diversity develop increased infiltration of activated CD4+ T-cell subsets.

Commented [SK3]: What you are describing next are local tumor immune responses, not systemic immune responses

DISCUSSION

The aim of this study was to identify independent gut microbial risk factors in cervical cancer patients receiving chemoradiation and to evaluate their impact on survival. We found BMI and gut diversity to be independent risk factors for RFS and OS in cervical cancer patients undergoing chemoradiation. The results indicate that overweight or obesity is a favorable prognostic factor independent of gut diversity. Additionally, our results demonstrate that patients with better clinical survival exhibit higher diversity as well as a distinct gut microbiome composition. Lastly the association between gut microbiome diversity and systematic immune signatures highlights helper CD4+ T cells as potential mediators of antitumor immunity upon CRT treatment.

Authors have previously described the gut microbiome and its effect on treatment outcomes for a variety of malignancies[39–41]. The diversity of gut microbiome is defined as the number and abundance distribution of distinct types of microorganisms colonizing within the gut[18]. In our study, higher alpha diversity at baseline correlated with an improved RFS and OS. High diversity implies more species harbor in the gut and suggests a difference in gut composition between short term and long term survivors. Our results imply that the diversity of gut microbiota

might be a shared benefit factor in those who respond well to CRT treatment. It is now generally accepted that the gut microbiome modulates immune responses, antitumor immunity, and clinical outcomes in a variety of malignancies[8,10,19]. The gut microbiome is thought to affect both innate and adaptive immune responses. Specifically how the gut microbiome exerts its influence continues to be explored, but this explanation may have important implications if specific taxa are found to change host response to treatment via immunomodulation[6]. In our study, T helper cell profiles at baseline correlate with gut diversity. These results confer that T cells and response to CRT are likely affected by the gut microbiota independent of other factors such as BMI. Using multi-color flow cytometry we performed correlation analysis on individual immune signatures and microbiota diversity. The frequency of helper CD4+ T cells were chiefly identified. Cervical cancer is considered to be an immunogenic tumor because its origin is dependent on a persistent infection with human papilloma virus (HPV), most often HPV16 or HPV18[20]. Previous studies have reported that the number and functional orientation of tumor-infiltrating CD4+ and CD8+ T cells and the presence of M1 type macrophages strongly correlates with survival in patients with cervical cancer after chemoradiation[20,21]. T cells are capable of rapid antigen-specific responses and play critical roles in immune recall responses. In addition to the percentage of CD4+ ~~t~~T cell subsets, the increase in CD4 Ki67, CD4 CD69, and CD4 PD1 in the patients with high microbiota diversity implies that gut microbiome also modulates the proliferation of certain immune cell populations. Recent studies have already reported that chemoradiotherapy for cervical cancer induces unfavorable immune changes reflected by a decreased number of circulating lymphocytes, both CD4+ and CD8+ T cells, and an increased percentage in myeloid-cell populations, including myeloid-derived suppressor cells and monocytes[20]. Whereas CD4+ T cells infiltrating in tumor microenvironment are thought to help the activity of other immune cells by releasing T cell

cytokines, circulating CD4+ T cell subsets reported here are probably inclined to reflect the role of gut microbiota on systemic immune responses. How peripheral memory CD4+ T cell signatures affect the efficacy of CRT treatment needs to be investigated in the future. Our study shows that the diversity of gut microbiota is associated with favorable response to CRT against cervical cancer. Considering the correlation between microbiota diversity and peripheral helper T cells being reshaped upon CRT treatment, we propose that patients with more diverse gut microbiota at baseline may benefit from CRT to a greater extent. This might be mediated by reprogramming systemic antitumor immune responses. The significance of our study lies in that the modulation of gut microbiota before treatment might provide an alternative way to enhance the efficacy of CRT, specifically in cases with positive lymph nodes and advanced stages in which systemic failure of current therapies represents a major challenge. Our results suggest that changes in the gut microenvironment contribute substantially to treatment success or failure, particularly in so-called immunogenic tumors like cervical cancer. Additionally, there is emerging data describing the influence of the gut microbiome as it pertains to radiotherapy[22]. Given that radiation can change the composition of the gut microbiome by altering the relative abundance of different taxa, we have to postulate whether it is these changes that ultimately alter the effectiveness of radiotherapy for cervical cancer[6,23,24].

Commented [SK4]: All the highlighted areas here discuss systemic immune responses, but you only show data for tumor immune responses

In our cohort, at baseline, a higher relative abundance of *Veillonella* resulted in a trend toward prolonged RFS and OS. Our own group has previously characterized the 16S rDNA fecal microbiome cervical cancer patients compared to healthy female controls, and have reported on differences in the relative abundance of specific taxa[42]. Our new findings support the hypothesis that organisms like *Veillonella* inhabiting the gut microbiome may be manipulated to improve cancer treatment response. Knowing specific gut microbial organisms that inhabit and undergo

changes in patients with cervical cancer during CRT provides further insight into mechanisms that may modulate immune response and potentiate treatment outcomes in cancer patients. The results of our study illustrate the potential of intentionally modifying the gut microbiota to accumulate CRT-tolerant species as an interventional strategy to enhance response of cervical cancer to CRT. Researchers have studied the treatment-enhancing utility of the gut microbiota in multiple areas of medicine[9,38]. For example, human fecal microbial transplants have protected germ-free mice from arsenic-induced mortality and reduced the number of antibiotic-resistant genes in patients with recurrent *Clostridium difficile* infections[39,40]. Also, Wang et al.[41] recently reported on the first case series of patients with immune checkpoint inhibitor-associated colitis successfully treated with fecal microbiota transplantation. With respect to how the gut microbiome can modulate the host response to chemotherapy, a previous review highlighted three important clinical elements: facilitation of drug efficacy, compromise of anticancer effects, and mediation of toxicity[42]. The authors went on to predict how the gut microbiome could be modified in clinical practice to increase cancer treatment efficacy and reduce toxicity. For example, in a murine model, radiation-induced dysbiosis increased the susceptibility of mice to radiotherapy-related gastrointestinal toxic effects[23]. Determining whether changes in the human gut microbiome during CRT affect patients' risk of treatment-related toxic effects may be an area for further investigation.

The "obesity paradox", which suggest a positive effects of increasing BMI as it pertains to a specific disease, was firstly reported in heart failure[25], but has since been described in a variety of disease processes including coronary artery disease, kidney disease, diabetes, and a variety of malignancies, including other gynecologic cancers[26–28]. Theories centered around the "obesity paradox" suggest that patients with a high BMI may be better able to withstand cancer-induced consumption and stress compared with patients with a low BMI[37]. Other theories include greater

metabolic reserve, an attenuated response to hormones involved in the renin–angiotensin–aldosterone system, fitness and its association with adiposity and clinical prognosis, and unmeasured confounding factors[26]. For example, in uterine cancer it has been reported that the risk of recurrence differed significantly by BMI[29]. Specifically, a greater proportion of obese women (BMI ≥ 40) met criteria for having a low risk of recurrence, while thin women tended to have a high intermediate risk or recurrence. There have been many studies investigating the impact of BMI on cervical cancer, but the association between BMI and cervical cancer remains unclear[30]. Most cervical cancer is caused by a persistent infection with a high risk human papillomavirus (HPV). However, it has been suggested that obesity may increase the risk of cervical cancer[31]. Contributing factors include poor screening and that body fat distribution hormonally influences the risk of glandular cervical carcinoma like adenocarcinoma of the cervix[32,33].

In contrast, Brinton et al reported that body weight was not an independent prognostic factor for squamous cell tumors, and a slight increased risk of adenocarcinoma, although this was not significant[34]. Tornberg et al. reported that there was not a significant relationship between being overweight and cervical cancer[35] and a review conducted in 2008 by Lane et al. did not report a relationship between cervical cancer and obesity siting a of a lack of evidence[36]. Finally, a meta-analysis done by Poorolajal et al. in 2016 indicated that being overweight (BMI 25–29.9 kg/m2), is not associated with an increased risk of cervical cancer, but that obesity (BMI ≥30 kg/m2) is weakly associated with an increased risk of cervical cancer[30]. However, the authors warned that more evidence, based on large prospective cohort studies, is required to provide conclusive evidence on whether or not BMI is associated with an increased risk of cervical cancer. These factors demonstrate the need to better understand if and how obesity increases cervical cancer risk. The inconsistent conclusions among studies investigating the association between

BMI and cervical cancer may be attributed to numerous factors including, but not limited to, patient selection criteria, sample size and generalizability of the study population to the general public. Among these factors, patient selection criteria may be especially important, because tumor histology seems to be closely associated with BMI.

The strengths of this study include the use of careful clinical staging, histopathology, and reliable phylogenetic and statistical analysis to assess bacterial community compositional changes using both microbial divergence and taxon-based methods. Additionally, we followed a complete protocol for 16S sequencing ranging from the sample collection method to DNA extraction and sequencing, thus limiting artifactual variations. Although this study yielded intriguing findings, it was limited by its small sample. Consequently, the sample size limited our ability to weigh statistical power. However, results presented herein provide solid evidence of the effect of CRT on the gut microbiome.

In conclusion, our study demonstrated that gut diversity is a significant factor for predicting OS in CC patients undergoing CRT when BMI is accounted for, and may help explain the "obesity paradox" in cancer response. Our study shows that the diversity of gut microbiota is associated with a favorable response to chemoradiation against cervical cancer. Considering the correlation between microbiota diversity and T cells being influenced with CRT treatment, patients with more diverse gut microbiota at baseline may benefit from CRT to a greater extent. The significance of our study lies in that the modulation of gut microbiota before CRT might provide an alternative way to enhance the efficacy of CRT but this needs to be validated in large cohort studies. Studies exploring the relationship between gut diversity, CRT, and treatment efficacy are needed to further understand the role of the gut microbiome in treatment outcomes.

ONLINE METHODS

Participants and clinical data. Gut microbiome and cervical swab samples were collected prospectively from cervical cancer patients according to a protocol approved by The University of Texas MD Anderson Cancer Center Institutional Review Board (MDACC 2014-0543) for patients with biopsy-proven carcinoma of the cervix treated at MD Anderson and the Lyndon B. Johnson Hospital Oncology Clinic from September 22, 2015, to January 11, 2019. All patients had new diagnoses of locally advanced, nonmetastatic carcinoma of the cervix and underwent definitive CRT with EBRT followed by brachytherapy. Patients received a minimum of 45 Gy via EBRT in 25 fractions over 5 weeks with weekly cisplatin followed by two brachytherapy sessions at approximately weeks 5 and 7 with EBRT in between for gross nodal disease or persistent disease in the parametrium. Patients with stage IB1 cancer were given CRT due to the presence of nodal disease. Clinical variables, demographics, and pathologic reports were abstracted from electronic medical records.

Sample collection and DNA extraction. Stool was collected from all patients by a clinician performing rectal exams at five time points (baseline; weeks 1, 3, and 5 of radiotherapy; and 3 months after CRT completion) using a matrix-designed quick-release Isohelix swab to characterize the diversity and composition of the microbiome over time. The swabs were stored in 20 µl of protease K and 400 µl of lysis buffer (Isohelix) and kept at -80°C within 1 h of sample collection.

16S rRNA gene sequencing and sequence data processing. 16S rRNA sequencing was performed for fecal samples obtained from all patients at four time points to characterize the diversity and composition of the microbiome over time. 16S rRNA gene sequencing was done at

the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine. 16S rRNA was sequenced using approaches adapted from those used for the Human Microbiome Project[43]. The 16S rDNA V4 region was amplified via polymerase chain reaction with primers that contained sequencing adapters and single-end barcodes, allowing for pooling and direct sequencing of polymerase chain reaction products. Amplicons were sequenced on the MiSeq platform (Illumina) using the 2 x 250-bp paired-end protocol, yielding paired-end reads that overlapped nearly completely. Sequence reads were demultiplexed, quality-filtered, and subsequently merged using the USEARCH sequence analysis tool (version 7.0.1090) (4). 16S rRNA gene sequences were bundled into operational taxonomic units at a similarity cutoff value of 97% using the UPARSE algorithm[44]. To generate taxonomies, operational taxonomic units were mapped to an enhanced version of the SILVA rRNA database containing the 16Sv4 region. A custom script was used to create an operational taxonomic unit table from the output files generated as described above for downstream analyses of α-diversity, β-diversity, and phylogenetic trends. Principal coordinates analysis was performed by institution and sample set to make certain no batch effects were present.

Flow Cytometry. Immunostaining was performed according to standard protocols. Cells were fixed using the Foxp3/Transcription Factor Staining Buffer Set (eBioscience, Waltham, MA) and stained with a 16 color panel with antibodies from Biolegend (San Diego, CA), BD Bioscience (San Jose, CA), eBioscience (Waltham, MA), and Life Technologies (Carlsbad, CA). Analysis was performed on a 5-laser, 18 color LSRFortessa X-20 Flow Cytometer (BD Biosciences, San Jose, CA). Analysis was performed using FlowJo version 10 (Flowjo LLC, Ashland, OR)~~on Flowjo Software (INFO)~~. Briefly, cells were stained with intracellular mAB for 30 minutes at 4C in the

Commented [SK5]: Site this reference here: Dorta-Estremera S, Colbert LE, Nookala SS, Yanamandra AV, Yang G, Delgado A, Mikkelson M, Eifel P, Jhingran A, Lilie LL, Welsh J, Schmeler K, ***Sastry JK***, Klopp A. Kinetics of intratumoral immune cell activation during chemoradiation for cervical cancer. Int J Radiat Oncol Biol Phys (2018) 102(3):593-600. PMID: 30017792.

presence of anti-Cd16/Cd32 mAB (BD Bioscience, San Jose, CA), fixed with Foxp3/Transcription Factor Staining Buffer Set (eBioscience), and held in FACS (Corning, 2 mM EDTA, 2% FBS). Counting beads (Thermo Fisher) were used for single color controls.

Commented [SK6]: In place of this highlighted area, use the details from the reference shown above or just site that reference for procedure details

Statistical analyses. For microbiome analysis, rarefaction depth was set at 7066 reads. The ISD index was used to evaluate α-diversity (within samples), and principle coordinates analysis of unweighted UniFrac distances was used to examine β-diversity (between samples). Patient and tumor characteristics were analyzed by univariate and multivariate Cox regression models for Recurrence-free survival (RFS) and Overall survival (OS) based on univariate p-value < 0.2. Characteristics included age, body mass index (BMI), race, stage, grade, histology, nodal status, smoking status, antibiotic use and max tumor size. For each outcome of interest, a multivariate Cox regression analysis was performed to adjust for the effects of prognostic factors identified on univariate analysis as influencing survival in cervical cancer. These analyses were conducted using covariates with $p \leq 0.2$ in a stepwise fashion. Alpha (within sample) diversity was evaluated using Shannon diversity index (SDI). The relative abundance of microbial taxa, classes, and genera was determined using LDA Effect Size[45], applying the one-against-all strategy with a threshold of 2 for the logarithmic LDA score for discriminative features and α of 0.05 for factorial Kruskal-Wallis testing among classes. LDA Effect Size analysis was restricted to bacteria present in 20% or more of the study population. Kaplan-Meier curves were generated for patients with normal BMI and overweight/obese BMI based on Cox analysis and clostridia abundance. The significance of differences was determined using the log-rank test. Gut microbial diversity, RFS, and OS were also compared using Pearson correlation, linear regression, and Cox regression analysis.

ACKNOWLEDGEMENTS

This work was supported in part by the Radiological Society of North America Resident/Fellow Award (to L.E.C.), the NIH/NCI under award number P30CA016672, and an NIH T32 grant (#5T32 CA101642-14; to T.T.S.). This study was partially funded by the MD Anderson HPV-Related Cancers Moon Shot program (to L.E.C. and A.K.). The human subjects who participated in this study are gratefully acknowledged.

AUTHOR CONTRIBUTION

All authors were involved in subject identification, data collection, interpretation of the statistical analysis, and review and approval of the final manuscript. The study concept was conceived by L.E.C., A.K., and T.T.S. The manuscript was written by T.T.S.

COMPETING INTERESTS

The authors report no conflicts of interest, financial or otherwise, related to the subject of this article.

ROLE OF FUNDING SOURCES

The funding sources were not involved in the development of the research hypothesis, study design, data analysis, or writing of the manuscript. Data access was limited to the authors of the manuscript.

References:

1. Islami, F., Fedewa, S. A. & Jemal, A. Trends in cervical cancer incidence rates by age, race/ethnicity, histological subtype, and stage at diagnosis in the United States. *Prev. Med.* **123**, 316–323 (2019).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA. Cancer J. Clin.* **69**, 7–34 (2019).
3. Monk, B. J., Tewari, K. S. & Koh, W.-J. Multimodality therapy for locally advanced cervical carcinoma: state of the art and future directions. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25**, 2952–2965 (2007).
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
5. Cho, J. A. & Chinnapen, D. J. F. Targeting friend and foe: Emerging therapeutics in the age of gut microbiome and disease. *J. Microbiol.* **56**, 183–188 (2018).
6. Helmink, B. A., Khan, M. A. W., Hermann, A., Gopalakrishnan, V. & Wargo, J. A. The microbiome, cancer, and cancer therapy. *Nat. Med.* **25**, 377–388 (2019).
7. Cerf-Bensussan, N. & Gaboriau-Routhiau, V. The immune system and the gut microbiota: friends or foes? *Nat. Rev. Immunol.* **10**, 735–744 (2010).
8. McQuade, J. L., Daniel, C. R., Helmink, B. A. & Wargo, J. A. Modulating the microbiome to improve therapeutic response in cancer. *Lancet Oncol.* **20**, e77–e91 (2019).
9. Roy, S. & Trinchieri, G. Microbiota: a key orchestrator of cancer therapy. *Nat. Rev. Cancer* **17**, 271–285 (2017).
10. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018).
11. Dewan, M. Z. *et al.* Fractionated but Not Single-Dose Radiotherapy Induces an Immune-Mediated Abscopal Effect when Combined with Anti-CTLA-4 Antibody. *Clin. Cancer Res.* **15**, 5379–5388 (2009).
12. Lugade, A. A. *et al.* Local Radiation Therapy of B16 Melanoma Tumors Increases the Generation of Tumor Antigen-Specific Effector Cells That Traffic to the Tumor. *J. Immunol.* **174**, 7516–7523 (2005).
13. Lhuillier, C., Rudqvist, N.-P., Elemento, O., Formenti, S. C. & Demaria, S. Radiation therapy and anti-tumor immunity: exposing immunogenic mutations to the immune system. *Genome Med.* **11**, 40 (2019).
14. Lee, Y. *et al.* Therapeutic effects of ablative radiation on local tumor require CD8+ T cells: changing strategies for cancer treatment. *Blood* **114**, 589–595 (2009).
15. Ngwa, W. *et al.* Using immunotherapy to boost the abscopal effect. *Nat. Rev. Cancer* **18**, 313–322 (2018).
16. Grassberger, C., Ellsworth, S. G., Wilks, M. Q., Keane, F. K. & Loeffler, J. S. Assessing the interactions between radiotherapy and antitumour immunity. *Nat. Rev. Clin. Oncol.* **16**, 729–745 (2019).
17. Sivan, A. *et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy. *Science* **350**, 1084–1089 (2015).
18. Jin, Y. *et al.* The Diversity of Gut Microbiome is Associated With Favorable Responses to Anti-Programmed Death 1 Immunotherapy in Chinese Patients With NSCLC. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **14**, 1378–1389 (2019).
19. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).

20. Meir, H. van *et al.* Impact of (chemo)radiotherapy on immune cell composition and function in cervical cancer patients. *OncoImmunology* **6**, e1267095 (2017).
21. Jordanova, E. S. *et al.* Human leukocyte antigen class I, MHC class I chain-related molecule A, and CD8+/regulatory T-cell ratio: which variable determines survival of cervical cancer patients? *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **14**, 2028–2035 (2008).
22. Muls, A. *et al.* Systematic Review: The Impact of Cancer Treatment on the Gut and Vaginal Microbiome in Women With a Gynecological Malignancy. *Int. J. Gynecol. Cancer* **27**, 1550–1559 (2017).
23. Gerassy-Vainberg, S. *et al.* Radiation induces proinflammatory dysbiosis: transmission of inflammatory susceptibility by host cytokine induction. *Gut* **67**, 97–107 (2018).
24. Sokol, H. & Adolph, T. E. The microbiota: an underestimated actor in radiation-induced lesions? *Gut* **67**, 1–2 (2018).
25. Lavie, C. J., Osman, A. F., Milani, R. V. & Mehra, M. R. Body composition and prognosis in chronic systolic heart failure: the obesity paradox. *Am. J. Cardiol.* **91**, 891–894 (2003).
26. Lavie, C. J., De Schutter, A. & Milani, R. V. Healthy obese versus unhealthy lean: the obesity paradox. *Nat. Rev. Endocrinol.* **11**, 55–62 (2015).
27. Walker, J. L. *et al.* Laparoscopy Compared With Laparotomy for Comprehensive Surgical Staging of Uterine Cancer: Gynecologic Oncology Group Study LAP2. *J. Clin. Oncol.* **27**, 5331–5336 (2009).
28. S, A.-A., A, S. & J, K. Obesity and cancer. *J. Med.* **33**, 73–86 (2002).
29. Gunderson, C. C., Java, J., Moore, K. N. & Walker, J. L. The impact of obesity on surgical staging, complications, and survival with uterine cancer: A Gynecologic Oncology Group LAP2 ancillary data study. *Gynecol. Oncol.* **133**, 23–27 (2014).
30. Poorolajal, J. & Jenabi, E. The association between BMI and cervical cancer risk: a meta-analysis. *Eur. J. Cancer Prev. Off. J. Eur. Cancer Prev. Organ. ECP* **25**, 232–238 (2016).
31. Nagaraju, G. P. *et al.* The role of adiponectin in obesity-associated female-specific carcinogenesis. *Cytokine Growth Factor Rev.* **31**, 37–48 (2016).
32. Martín-López, R. *et al.* Breast and cervical cancer screening in Spain and predictors of adherence. *Eur. J. Cancer Prev.* **19**, 239–245 (2010).
33. Lacey, J. V. *et al.* Obesity as a potential risk factor for adenocarcinomas and squamous cell carcinomas of the uterine cervix. *Cancer* **98**, 814–821 (2003).
34. Brinton, L. A. *et al.* Risk Factors for Cervical Cancer by Histology. *Gynecol. Oncol.* **51**, 301–306 (1993).
35. Törnberg, S. A. & Carstensen, J. M. Relationship between Quetelet's index and cancer of breast and female genital tract in 47,000 women followed for 25 years. *Br. J. Cancer* **69**, 358–361 (1994).
36. Lane, G. Obesity and gynaecological cancer. *Menopause Int.* **14**, 33–37 (2008).
37. Chen, H.-N. *et al.* The Impact of Body Mass Index on the Surgical Outcomes of Patients With Gastric Cancer: A 10-Year, Single-Institution Cohort Study. *Medicine (Baltimore)* **94**, e1769 (2015).
38. Kohl, K. D., Weiss, R. B., Cox, J., Dale, C. & Denise Dearing, M. Gut microbes of mammalian herbivores facilitate intake of plant toxins. *Ecol. Lett.* **17**, 1238–1246 (2014).
39. Coryell, M., McAlpine, M., Pinkham, N. V., McDermott, T. R. & Walk, S. T. The gut microbiome is required for full protection against acute arsenic toxicity in mouse models. *Nat. Commun.* **9**, 5424 (2018).

40. Millan, B. *et al.* Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent *Clostridium difficile* Infection. *Clin. Infect. Dis.* **62**, 1479–1486 (2016).
41. Wang, Y. *et al.* Fecal microbiota transplantation for refractory immune checkpoint inhibitor-associated colitis. *Nat. Med.* **24**, 1804–1808 (2018).
42. Alexander, J. L. *et al.* Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 356–365 (2017).
43. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
44. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
45. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

509

Date : 4/19/2020 4:31:15 PM
From : "Schmeler,Kathleen M" KSchmele@mdanderson.org
To : "Sims,Travis T." TTSims@mdanderson.org, "Colbert,Lauren Elizabeth" LColbert@mdanderson.org, "Karpinets,Tatiana V" TVKarpinets@mdanderson.org, "Ning,Matthew Stephen" MSNing@mdanderson.org, "El Alam,Molly B" MBEl@mdanderson.org, "Court,Kyoko" KCourt1@mdanderson.org, "Wu,Xiaogang" XWu10@mdanderson.org, "Mezzari, Melissa" [redacted] "Delgado Medrano,Andrea Yizel" AYDelgado@mdanderson.org, "Ajami,Nadim J" NAjami@mdanderson.org, "Solley,Travis N" TNSolley@mdanderson.org, "Ahmed-Kaddar,Mustapha" MAhmed10@mdanderson.org, [redacted]
Cc : "Klopp,Ann H" AKlopp@mdanderson.org, "Biegert, Greyson" Greyson.Biegert@uth.tmc.edu
Subject : Re: Manuscript - Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients
Attachment : Manuscript - Cervical Tumor Microbiome Botswana 4-14-20.KMS.docx;Botswana Manuscript - Table 1.[1].KMS.docx;Botswana Manuscript - Table 2.KMS.docx;

Hi all. See attached suggested edits

Thanks for including me!
Kathleen

--
Kathleen M. Schmeler, MD
Professor
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center

Phone: 713-745-3518
Fax: 713-792-7586

Mailing Address:
Unit 1362
PO Box 301439
Houston, TX 77230-1429

From: "Sims,Travis T." <TTSims@mdanderson.org>
Date: Tuesday, April 14, 2020 at 1:12 PM
To: Lauren Colbert <LColbert@mdanderson.org>, "Karpinets,Tatiana V" <TVKarpinets@mdanderson.org>, "Ning,Matthew Stephen" <MSNing@mdanderson.org>, "El Alam,Molly B" <MBEl@mdanderson.org>, "Court,Kyoko" <KCourt1@mdanderson.org>, "Wu,Xiaogang" <XWu10@mdanderson.org>, "Mezzari,

Melissa" [redacted], "Delgado Medrano,Andrea Yizel" <AYDelgado@mdanderson.org>, Nadim Ajami <NAjami@mdanderson.org>, "Solley,Travis N" <TNSolley@mdanderson.org>, "Ahmed-Kaddar,Mustapha" <MAhmed10@mdanderson.org>, "joseph.petrosino@bcm.edu" [redacted], Kathleen Schmeler <KSchmele@mdanderson.org>, Surbhi Grover [redacted]

Cc: Ann Klopp <AKlopp@mdanderson.org>, "Biegert, Greyson" <Greyson.Biegert@uth.tmc.edu>

Subject: Manuscript - Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients

Hello all,

We have completed the first draft of the manuscript for our project "Tumor Microbial Diversity and Compositional Differences in Botswana Cervical Dysplasia and Cervical Cancer Patients".

You have been included on the attached manuscript given your participation and clinical interest in this subject area. We will be submitting this manuscript to The *International Journal of Gynecological Cancer* (*IJGC*).

Attached you will find the manuscript, tables, and figures. Please let me know if you have any questions or concerns, or any edits to the manuscript. Lastly, let us know if you identify any other authors you feel should be included.

I am grateful for your time and feedback regarding this project! We hope to submit by 4/24/20.

Best,
Travis

Travis T. Sims, MD, MPH
Fellow
Department of Gynecologic Oncology & Reproductive Medicine
The University of Texas MD Anderson Cancer Center
ttsims@mdanderson.org
C [redacted]
T 346-315-9781
P 713-404-6828

Title: Tumor Microbial Diversity and Compositional Differences ~~in~~ Among Women in Botswana with High-Grade Cervical Dysplasia and Cervical Cancer ~~Patients~~

Authors and Affiliations

Travis T. Sims, MD, MPH[1*], Greyson W.G. Biegert, BS[2*], Kebatshabile Ngoni, BSc[3], Molly B. El Alam, MPH[2], Matthew S. Ning, MD, MPH[2], Tatiana V. Karpinets, PhD[4], Kyoko Yoshida-Court, PhD[2], Xiaogang Wu, PhD[4], Andrea Y. Delgado Medrano, BS[2], Melissa P. Mezzari, PhD[5], Joseph Petrosino, PhD[5], Nadim J. Ajami, PhD[4], Travis Solley, BS[2], Mustljapha Ahmed-Kaddar, BS[2], Kathleen M. Schmeler, MD[1], Lauren E. Colbert, MD, MSCR[2], Stephen Hahn, MD[2], Ann H. Klopp, MD, PhD[2], Surbhi Grover, MD, MPH[6]

[1]Department of Gynecologic Oncology and Reproductive Medicine, Division of Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX

[2]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

[3]Department of Biological Sciences, University of Botswana, Gaborone, Botswana

[4]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX

[5]Department of Molecular Virology and Microbiology, Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX

[6]Department of Radiation Oncology, Botswana-University of Pennsylvania Partnership, Gaborone, Botswana, and Philadelphia, PA

*These authors contributed equally.

Corresponding Author

Travis T. Sims, MD, MPH

Department of Gynecologic Oncology and Reproductive Medicine, Division of Surgery

The University of Texas MD Anderson Cancer Center

1155 Herman Pressler, CPB 6.3258, Unit 1362

Houston, TX 77030

Phone: 713-563-4578

Fax: 713-745-2398

Email: ttsims@mdanderson.org

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Research Support

This research was supported in part by the National Institutes of Health (NIH) through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant 5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK).

Role of Funding Sources

47 The funding sources were not involved in the research hypothesis development, study design,
48 data analysis, or manuscript writing. Data access was limited to the authors of this manuscript.

ABSTRACT

Introduction We characterized the cervical 16S rDNA microbiome of high-grade cervical dysplasia and locally advanced cervical cancer in patients in Botswana. **Methods** Our prospective study included 31 patients (21 with dysplasia and 10 with cancer). We used the Shannon diversity index to evaluate alpha (within sample) diversity and UniFrac (weighted and unweighted) and Bray-Curtis distances to evaluate beta (between sample) diversity. We compared the relative abundance of microbial taxa between samples using linear discriminant analysis effect size. **Results** Alpha diversity was significantly higher in patients with cervical cancer ~~patients~~ than in patients cervical dysplasia ~~patients~~ ($p<0.05$). Beta diversity (weighted UniFrac Bray-Curtis, $p<0.01$) also significantly differed. The results of linear discriminant analysis effect size demonstrated that multiple taxa significantly differed between patients with cervical dysplasia ~~and~~ vs. cancer ~~patients~~. *Lachnospira* bacteria, in the *Clostridia* class, were significantly enriched in patients with cervical dysplasia ~~patients~~, while *Proteobacteria*, members of the *Firmicutes* phyla and the *Comamonadaceae* family were enriched in patients with cervical cancer ~~patients~~. **Discussion** The results of our study suggest that differences exist in the diversity and composition of the cervical microbiota between patients with cervical dysplasia and patients with cervical cancer ~~patients~~ in Botswana. Additional studies are needed to validate these findings in larger cohorts to determine the biological significance of these observed differences in women living in Botswana as well as ~~southern Africa~~other regions of the world.

Commented [MOU1]: Do you have comparison data for CIN and cancer pts in the US or elsewhere.... I think reviewers will want to know how the Botswana data compares to the US and other regions. If not available – add this to the limitations section of the Discussion

Keywords: Cervical dysplasia; cervical cancer; gynecologic cancer; cervical microbiota; microbiome; HIV; Botswana

Highlights:

- In this cohort of women in Botswana, cervical microbiome diversity was higher in women with cervical cancer ~~patients than in~~compared with cervical dysplasia ~~patients~~.
- The cervical microbiota of women with cervical cancer have a distinct composition compared with those of women with cervical dysplasia.
- Currently, there ~~is an important gap in the number of~~are limited studies investigating the cervical microbiome and gynecologic cancers in women in sub-Saharan Africa~~n patients~~.

INTRODUCTION

Cervical cancer is one of the most common malignancies globally and the most common cause of cancer death among African women[1]. More than half a million new cases of invasive cervical cancer are expected to be diagnosed worldwide in 2020, resulting in more than 300,000 deaths[2]. African women have a far higher risk of cervical cancer than do women in regions with more access to preventative health care screening[1]. Fourteen percent of the world's cervical cancer cases and 18% of cervical cancer-related deaths occur in women living in sub-Saharan Africa[1,3]. The incidence of cervical cancer in southern Africa, which includes the countries of Botswana, Lesotho, Namibia, South Africa, and Swaziland, is expected to increase by roughly 35% by 2030[1].

It is well established that persistent exposure to the human papilloma virus (HPV) is an antecedent to cervical cancer[4]. Women living with human immunodeficiency virus (HIV) are at increased risk of persistent HPV infection and ultimately, cervical cancer, despite access to anti-retroviral therapy[5]. The high regional prevalence of HIV in countries such as Botswana underscores the importance of cervical cancer prevention in these regions. Botswana established one of the original nationwide HIV treatment programs[6] in Africa, but despite a corresponding decline in HIV-associated mortality, the incidence of cervical cancer remains among the highest globally (36.6 per 100,000), with nearly two-thirds of cases occurring in HIV-positive women[7].

The microbiome has recently been demonstrated to play a critical role in cancer progression and metastasis and cancer-directed therapy response[8]. The female cervix is a microbiome-rich environment, but the effect of this microbiome on cervical dysplasia and progression to cervical cancer ~~development~~ and progression is limited and not well understood[9].

Commented [MOU2]: You may want to say sub-Saharan Africa or be specific to Botswana. Cervical cancer is relatively rare in many countries in Northern Africa

Commented [MOU3]: I would stick to sub-Saharan Africa (SSA) – it is confusing to jump between Southern Africa and SSA. And Southern Africa is defined differently by different groups – if you want to use "Southern Africa" then I would specify that you are using the UN designation... as others include Mozambique and additional countries

Given the expected incidence of cervical cancer in 2020, understanding the effect of the cervical flora on cancer progression and response, as well as the converse effect of treatments such as chemoradiation therapy, represents a critical unmet need, especially in vulnerable populations, such as women residing in Botswana.

To our knowledge, no published studies exist that specifically explore the cervical tumor microbiome in women in Botswana. Cervical cancer is uniquely positioned for such a crucial investigation, as it allows direct visualization and contact with the primary tumor at the initiation of treatment.

Because cervical microbial differences can affect cervical cancer risk and treatment through several pathways, we characterized the 16S rDNA cervical microbiome of women with cervical dysplasia and locally advanced cervical cancer in Botswana. We hypothesize that the cervical microbiome of patients with cervical cancer ~~patients~~ is distinct from that of patients with dysplasia ~~patients~~. We theorize that the longitudinal identification of persistent bacterial strains that are associated with the cervical microbiome will allow us to further study the organisms that stably colonize cervical cancers, detect bacterial strains that are associated with treatment response, and lay the groundwork for developing interventions that alter the tumor microbiota to improve cancer outcomes.

PATIENTS AND METHODS

Participants and Clinical Data

We prospectively identified patients with newly diagnosed, biopsy-proven high-grade cervical dysplasia or locally advanced, non-metastatic cervical carcinoma who were treated at the

University of Botswana General Hospital oncology clinic between July 24, 2018, and February 22, 2019. ~~The study protocol, the final approved informed consent document and the subject recruitment information were submitted to the~~ Institutional Review Board (IRB) ~~and samples used for this study were obtained following ethical approval by the IRB at~~approval for the study was obtained the University of Botswana (~~[IRB reference number:~~ UBR/RES/IRB/BIO/045~~]~~), the University of Pennsylvania (~~IRB reference number:~~ 830039), and T~~t~~he University of Texas MD Anderson Cancer Center (~~IRB reference number:~~ MDACC 2014-0543). The subject's informed consent was mandatory for study participation and was obtained in writing.

Patient ineligibility criteria included incident or prevalent cancer other than cervical cancer and currently pregnant women. Medical history and current medication use were assessed via an in-person interview with a clinical provider or trained study staff. We reviewed patients' medical records to obtain demographic and clinico-pathologic data. All cancer patients had a new diagnosis of locally advanced, non-metastatic carcinoma of the cervix ~~and underwent~~with planned definitive chemoradiation (CRT) with external beam radiation therapy followed by brachytherapy~~, but~~. All study samples ~~used for this st~~udy were collected prior to any cancer therapy.

Commented [MOU4]: How about for the dysplasia patients? Pre-treatment as well?

Sample Collection and DNA Extraction

Cervical samples were collected using a matrix-designed quick-release Isohelix swab. The swabs were placed in 20 µL of protease K and 400 µL of lysis buffer (Isohelix) and stored at –80°C within 1 hour of sample collection. Bacterial genomic DNA was extracted using a MO BIO

PowerSoil DNA Isolation Kit (MO BIO Laboratories). Samples were shipped to the US for downstream applications that include DNA processing and sequencing.

16S rRNA Gene Sequencing and Sequence Data Processing

16S rRNA gene sequencing of the cervical swabs was performed at the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine (Houston, Texas, USA) using methods adapted from those used for the Human Microbiome Project.[10] The 16S rDNA V4 region was amplified by PCR using primers that contained sequencing adapters and single-end barcodes, allowing the pooling and direct sequencing of PCR products. Amplicons were sequenced on the MiSeq platform (Illumina) using the 2x250-bp paired-end protocol, yielding paired-end reads that overlapped almost completely. The sequence reads were de-multiplexed, quality filtered, and subsequently merged using USEARCH version 7.0.1090 (4). 16S rRNA gene sequences were clustered into OTUs at a similarity cut-off value of 97% using the UPARSE algorithm.[11] To generate taxonomies, we mapped OTUs to an optimized version of the SILVA rRNA database containing the 16S v4 region. A custom script was used to construct an OTU table from the output files generated, as described above, for downstream analyses of alpha diversity, beta diversity, and phylogenetic trends. Principal coordinates analysis was performed by institution and sample set to ensure that no batch effects were present.

Statistical Analyses

For the microbiome analysis, the rarefaction depth was set at 3651 reads. Alpha (within sample) diversity was examined using the Shannon diversity index, and beta (between sample) diversity was examined using UniFrac (weighted and unweighted) and Bray-Curtis distances. We compared the relative abundance of microbial taxa and genera between samples; we then determined differentially abundant bacterial genera by case status using linear discriminant analysis (LDA) effect size (LEfSe),[12] applying the 1-against-all strategy with a threshold of 4 on the logarithmic LDA score for discriminative features and an α of 0.05 for the factorial Kruskal-Wallis test among classes. LEfSe was restricted to bacteria that were present in 20% or more of the study population. Observed differences were subjected to paired analysis using two sample Z test for proportions, or Student t test where appropriate.

RESULTS

We characterized the 16S rDNA cervical microbiome in 31 patients with cervical dysplasia (n=21) and cancer ~~patients (21 with dysplasia and~~(n=10~~ with cancer~~). Clinico-pathologic data for all patients are summarized in Table 1. Cervical dysplasia ~~patients were~~was classified according to ~~their~~ the histological grade of cervical intraepithelial neoplasia ([CIN] ~~stage I III~~1-3). 18 (~~Approximately~~ 58%) ~~of the~~ patients ~~in the study (18 of 31)~~ had CIN 2 ~~stage III~~, 3 (x%) had CIN 3 and 10 (~~approximately~~ 32%) ~~(10 of 31)~~ had cervical cancer (in all cases, squamous cell cancer with moderate or poor differentiation). HPV status was unknown in all patients at the time of cervical sampling.

Commented [MOU5]: We don't use stage for CIN

We first analyzed patients' microbiota with respect to HIV status. Neither α diversity (p=0.8) nor β diversity (p=0.19) varied by HIV status (Figure 1A,B), and the top 10 most

abundant genera were similar among all cervical cancer patients (Figure 1C), suggesting that bacterial taxa dominance ~~does~~ may not vary by HIV status.

We then sought to extend our analysis to characterize variations in the cervical microbiome by cervical dysplasia vs cervical cancer. Patients' clinical and demographic characteristics are displayed in Table 2. The mean age and BMI were similar between patients with cervical dysplasia ~~patients~~ vs. ~~and~~ cervical cancer ~~patients~~ (mean age, 41.8 vs 50.7 years [p=0.1], and mean BMI, 26.3 vs. 30.0 kg/m^2 [p=0.19], respectively). We observed a statistically significant higher α diversity, as measured by SDI ($p<0.05$), in cervical dysplasia patients than in cervical cancer patients (Figure 2A). Patients with CIN ~~III~~ 3 ~~patients~~ tended to have higher α diversity than ~~did~~ those with CIN ~~II patients~~2 (Figure 2B). As with α diversity, overall β diversity differed significantly by cancer status (weighted Bray-Curtis Unifrac; $p<0.01$) (Figure 2C,D). The top 10 most abundant genera in cervical samples were similar among all patients with cervical dysplasia and cervical cancer ~~patients~~ (Figure 2E). The percentage of subjects with a cervical microbiome dominated by *Lactobacillus* was low in both groups but lower in the cervical cancer cohort (1 of 10 patients).

We used LEfSe to identify the bacterial genera that were differentially enriched in our cohort of patients ($p<0.05$, LDA score >2). We found that the genera *Ersipelotrichia*, *Erysipelotrichales*, *Erysipelotrichaceae*, and *Ruminiclostridium* were enriched in HIV-positive patients, while only *Filifactor* was significantly enriched in HIV-negative patients (Figure 1D,E). We found that the genus *Lachnospira*, in the *Clostridia* class of bacteria, was significantly enriched in cervical dysplasia patients, while several *Proteobacteria* taxa (*Betaproteobacteria*, *Gammaproteobacteria*, and *Burkholderiaceae*) and members of the *Firmicutes* phyla

Commented [MOU6]: I would move this statement to the discussion and report findings only in the results section

Commented [MOU7]: Are the numbers high enough to make this statement

(*Erysiopelotrichaceae* and *Synergistaceae*) and the *Comamonadaceae* family were significantly enriched in cervical cancer patients ($p<0.05$, LDA score >2) (Figure 2F,G).

DISCUSSION

In this study, we characterized the cervical microbiome of women with cervical dysplasia and cervical cancer ~~patients~~ living in Botswana. We hypothesized that the cervical microbiome ~~of cervical cancer patients would be distinct from that of dysplasia patients~~would differ between the two groups. We observed significant differences in cervical α and β diversity between these groups of patients, as well as compositional differences. The results of an overall analysis of α and β diversity revealed that the groups did not differ in regard to HIV status.

The influence of the cervical cancer microbiome site throughout treatment is poorly understood. Research has focused on exploring the relative abundance of bacteria in the vaginal epithelium, with the assignment of community-state types based on the richness of *Lactobacilli* species[13-15]. The presence and abundance of specific *Lactobacilli* species, for example, *L. crispatus*, *L. gasseri*, or *L. jensenii*, is thought to be associated with a predisposition to bacterial vaginosis (BV) and other pro-inflammatory states[16,17].

However, despite the comparative wealth of data focused on the vaginal microbiome, the ectocervical microbiome has yet to be well described. Most studies have concentrated on characterizing it in the setting of pregnancy or pelvic inflammatory disease. Previous studies using 16S rDaNA sequencing have suggested that in pregnancy, cervical microbiota diversity differs by race[18] and that the presence of non-*Lactobacillus* community state types is associated with a robust cervical inflammatory response in the setting of pre-term, premature membrane

243 rupture[19,20]. Wang et al. demonstrated that in patients with pelvic inflammatory disease, the
244 cervical microbiota is dominated by *Lactobacillus* and *Gardnerella*, again suggesting that the
245 abundance of these different taxa is associated with both acute and chronic inflammatory states[21].
246 It is thought that these states of polybacterial dysbiosis and chronic local inflammation
247 encourage the perseverance of HPV, which ultimately promotes the development of cervical
248 dysplasia and carcinogenesis in the setting of persistent HPV exposure[15,17,22-25].
249 Persistent HPV infections are thought to trigger an innate immune response, resulting in
250 the suppression of infected cervicovaginal mucosal cells[16,26,27]. An altered mucosal
251 microenvironment leads to the growth of anaerobic organisms at the expense of *Lactobacillus*
252 growth, creating cervicovaginal dysbiosis[28]. LEfSe was designed to detect bacterial taxa that are
253 associated with a specific state[29]. In our study, LEfSe identified *Clostridia*, *Firmicutes*, and
254 *Lachnospira* as taxa that were negatively associated with cervical cancer and several
255 *Proteobacteria* as taxa that were positively associated with cervical cancer compared with
256 cervical dysplasia.
257 Dysbiosis causes cervicovaginal inflammation and other unfavorable changes in the
258 cervicovaginal mucosal barrier. Worldwide, the most common type of cervicovaginal dysbiosis,
259 which is defined as a cervicovaginal microbiome that is not dominated by *Lactobacilli*, is BV[30].
260 BV is characterized by a persistent decrease in *Lactobacilli* and an increase in fastidious
261 anaerobes[26]. Globally, the prevalence of BV is highest in women living in sub-Saharan Africa
262 and in women of sub-Saharan African descent[30]. Cervicovaginal dysbiotic states, such as BV,
263 lead to an altered metabolic profile and reduced cervicovaginal barrier function. This dysbiotic
264 state is not only associated with an increased acquisition of HIV, but also with high-risk HPV,
265 cervical dysplasia, and ultimately cervical cancer[26,31]. The percentage of subjects with their

cervical microbiome dominated by *Lactobacillus* was low in our cohort of patients. The proportion of dysplasia patients with *Lactobacillus*-dominated cervical microbiomes was higher than that of cancer patients. The lack of *Lactobacilli* identified in our cervical dysplasia and cervical cancer patients supports this rationale and suggests that cervicovaginal microbes are important in preventing or enhancing the acquisition and pathogenesis of HPV and HIV. Identifying the microbes that are associated with enhanced pathogenesis and ultimately oncogenesis or tumorigenesis is especially important in susceptible populations such as HIV-positive women in Botswana. Historically, microbiome cervical cancer research has been limited to mainly Western industrialized populations. We hope that our findings in women in Botswana provide a timely and critical glimpse into this uniquely vulnerable population.

The gut microbiome and its influence on carcinogenesis and prognosis has been well described, most notably in melanoma and colorectal cancer[8,32,33]. Bullman et al. recently identified colonization by *Fusobacterium* and its associated microbiome *Bacteriodes*, *Selenomas*, and *Prevotella* at both the primary tumor and the distant paired metastatic site in colorectal cancer. Thus, it is possible that the colonized organisms that inhabit the primary tumor migrate with primary tumor cells to distant locations and manipulate microbiota diversity at sites, ultimately leading to poor anti-tumor immunity[34]. Identifying the specific organisms that colonize the tumor microbiota will provide further insight into the mechanisms that modulate immune response and potentiate tumor cell growth[31].

Although the present study yielded intriguing findings, it was limited by its small sample size. We acknowledge this possible limitation, but our sample size is suggestive of the complexity associated with using 16S rDNA next-generation sequencing to evaluate the cervical microbiome in a remote population; complete data collection was limited, and field

Commented [MOU8]: An important limitation is the lack of a control group of "normal" patients without dysplasia or cancer. In addition, no comparison group from the US or elsewhere

circumstances were challenging. Our study design also prevents us from determining the causal associations or mechanisms that are associated with differences in the cervical microbiota and cervical dysplasia or cancer; this is an area that deserves further study. These limitations are unlikely to fully explain the large differences that we observed between cervical dysplasia and cancer patients.

In conclusion, our study demonstrated hypothesis-generating differences in the cervical microbial profiles of women in Botswana with cervical cancer ~~patients~~ compared to those ~~of~~ with cervical dysplasia ~~patients~~. The lack of *Lactobacilli* in our samples supports the rationale that cervicovaginal dysbiotic states, which are characterized by a persistent decrease in *Lactobacilli*, are associated with a higher incidence of HIV, cervical dysplasia, and cervical cancer. We anticipate that our findings will help improve our understanding of the essential functional role of the tumor microbiome in cervical cancer. Additional studies are needed to validate these findings in larger cohorts and to determine the biological significance of these observed differences in women living in southern Africa.

Conflicts of Interest

The authors report no conflicts of interest, financial or otherwise, related to the subject matter of the article submitted.

Author Contributions

All authors were involved with subject identification and data collection, interpretation of the statistical analysis, and review and approval of the final manuscript. The study concept was developed by LEC, AK, GWGB, and TTS. GWGB and TTS helped draft the manuscript.

Acknowledgements

This work was supported in part by the National Institutes of Health through MD Anderson's Cancer Center Support Grant P30 CA016672 and the National Institutes of Health T32 grant #5T32 CA101642-14 (TTS). This study was partially funded by the MD Anderson HPV-Related Cancers Moonshot (AK). We gratefully acknowledge the patients who participated in this study.

REFERENCES

1. De Vuyst H, Alemany L, Lacey C, et al. The burden of human papillomavirus infections and related diseases in sub-saharan Africa. *Vaccine*. 2013;31 Suppl 5:F32-46. doi:10.1016/j.vaccine.2012.07.092

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34. doi:10.3322/caac.21551

3. Klein C, Kahesa C, Mwaiselage J, West JT, Wood C, Angeletti PC. How the Cervical Microbiota Contributes to Cervical Cancer Risk in Sub-Saharan Africa. *Front Cell Infect Microbiol*. 2020;10. doi:10.3389/fcimb.2020.00023

4. Hammer A, Rositch A, Qeadan F, Gravitt PE, Blaakaer J. Age-specific prevalence of HPV16/18 genotypes in cervical cancer: A systematic review and meta-analysis. *Int J Cancer*. 2016;138(12):2795-2803. doi:10.1002/ijc.29959

5. Dryden-Peterson S, Bvochora-Nsingo M, Suneja G, et al. HIV Infection and Survival Among Women With Cervical Cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2016;34(31):3749-3757. doi:10.1200/JCO.2016.67.9613

6. Gaolathe T, Wirth KE, Holme MP, et al. Botswana's progress toward achieving the 2020 UNAIDS 90-90-90 antiretroviral therapy and virological suppression goals: a population-based survey. *Lancet HIV*. 2016;3(5):e221-230. doi:10.1016/S2352-3018(16)00037-0

7. Dryden-Peterson S, Medhin H, Kebabonye-Pusoentsi M, et al. Correction: Cancer Incidence following Expansion of HIV Treatment in Botswana. *PloS One*. 2015;10(9):e0138742. doi:10.1371/journal.pone.0138742

8. Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy. *Cancer Cell*. 2018;33(4):570-580. doi:10.1016/j.ccell.2018.03.015

9. Klein C, Gonzalez D, Samwel K, et al. Relationship between the Cervical Microbiome, HIV Status, and Precancerous Lesions. *mBio*. 2019;10(1). doi:10.1128/mBio.02785-18

10. Bombardelli C, Ayuso JH, Pelayo RG. A framework for human microbiome research. *Nature*. 2012;486(7402):215-221. doi:10.1038/nature11209

11. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10(10):996-998. doi:10.1038/nmeth.2604

12. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12(6). doi:10.1186/gb-2011-12-6-r60

13. Kyrgiou M, Mitra A, Moscicki A-B. Does the vaginal microbiota play a role in the development of cervical cancer? *Transl Res J Lab Clin Med*. 2017;179:168-182. doi:10.1016/j.trsl.2016.07.004

14. Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A*. 2011;108 Suppl 1:4680-4687. doi:10.1073/pnas.1002611107

15. Brotman RM, Shardell MD, Gajer P, et al. Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *J Infect Dis*. 2014;210(11):1723-1733. doi:10.1093/infdis/jiu330

16. van de Wijgert JHHM, Borgdorff H, Verhelst R, et al. The vaginal microbiota: what have we learned after a decade of molecular characterization? *PloS One*. 2014;9(8):e105998. doi:10.1371/journal.pone.0105998

17. King CC, Jamieson DJ, Wiener J, et al. Bacterial vaginosis and the natural history of human papillomavirus. *Infect Dis Obstet Gynecol*. 2011;2011:319460. doi:10.1155/2011/319460

18. Wheeler S, Pryor K, Antczak B, Truong T, Murtha A, Seed P. The relationship of cervical microbiota diversity with race and disparities in preterm birth. *J Neonatal-Perinat Med*. 2018;11(3):305-310. doi:10.3233/NPM-17111

19. Kacerovsky M, Pliskova L, Bolehovska R, et al. Lactobacilli-dominated cervical microbiota in women with preterm prelabor rupture of membranes. *Pediatr Res*. December 2019. doi:10.1038/s41390-019-0692-1

20. Kacerovsky M, Vrbacky F, Kutova R, et al. Cervical microbiota in women with preterm prelabor rupture of membranes. *PloS One*. 2015;10(5):e0126884. doi:10.1371/journal.pone.0126884

21. Wang Y, Zhang Y, Zhang Q, Chen H, Feng Y. Characterization of pelvic and cervical microbiotas from patients with pelvic inflammatory disease. *J Med Microbiol*. 2018;67(10):1519-1526. doi:10.1099/jmm.0.000821

22. Marullo R, Werner E, Zhang H, Chen GZ, Shin DM, Doetsch PW. HPV16 E6 and E7 proteins induce a chronic oxidative stress response via NOX2 that causes genomic instability and increased susceptibility to DNA damage in head and neck cancer cells. *Carcinogenesis*. 2015;36(11):1397-1406. doi:10.1093/carcin/bgv126

23. Mitra A, MacIntyre DA, Lee YS, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci Rep*. 2015;5:16865. doi:10.1038/srep16865

24. Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet Lond Engl*. 2001;357(9255):539-545. doi:10.1016/S0140-6736(00)04046-0

25. Balkwill FR, Mantovani A. Cancer-related inflammation: common themes and therapeutic opportunities. *Semin Cancer Biol*. 2012;22(1):33-40. doi:10.1016/j.semcancer.2011.12.005

26. van de Wijgert JHHM, Gill AC, Chikandiwa A, et al. Human papillomavirus infection and cervical dysplasia in HIV-positive women: potential role of the vaginal microbiota. *AIDS*. 2020;34(1):115–125. doi:10.1097/QAD.0000000000002381

27. Karim R, Tummers B, Meyers C, et al. Human papillomavirus (HPV) upregulates the cellular deubiquitinase UCHL1 to suppress the keratinocyte's innate immune response. *PLoS Pathog*. 2013;9(5):e1003384. doi:10.1371/journal.ppat.1003384

28. Ilhan ZE, Łaniewski P, Thomas N, Roe DJ, Chase DM, Herbst-Kralovetz MM. Deciphering the complex interplay between microbiota, HPV, inflammation and cancer through cervicovaginal metabolic profiling. *EBioMedicine*. 2019;44:675-690. doi:10.1016/j.ebiom.2019.04.028

29. Brooks JP, Edwards DJ, Blithe DL, et al. Effects of combined oral contraceptives, depot medroxyprogesterone acetate and the levonorgestrel-releasing intrauterine system on the vaginal microbiome. *Contraception*. 2017;95(4):405-413. doi:10.1016/j.contraception.2016.11.006

30. van de Wijgert JHHM, Jespers V. The global health impact of vaginal dysbiosis. *Res Microbiol*. 2017;168(9-10):859-864. doi:10.1016/j resmic.2017.02.003

31. Borgdorff H, Gautam R, Armstrong SD, et al. Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. *Mucosal Immunol*. 2016;9(3):621-633. doi:10.1038/mi.2015.86

32. Wu S, Rhee K-J, Albesiano E, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med*. 2009;15(9):1016-1022. doi:10.1038/nm.2015

33. Arthur JC, Perez-Chanona E, Mühlbauer M, et al. Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science*. 2012;338(6103):120-123. doi:10.1126/science.1224820

34. Bullman S, Pedamallu CS, Sicinska E, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science*. 2017;358(6369):1443-1448. doi:10.1126/science.aal5240

35. Yang Y, Weng W, Peng J, et al. Fusobacterium nucleatum Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor-κB, and Up-regulating Expression of MicroRNA-21. *Gastroenterology*. 2017;152(4):851-866.e24. doi:10.1053/j.gastro.2016.11.018

36. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe*. 2013;14(2):195-206. doi:10.1016/j.chom.2013.07.012

Table Legends

Table 1. Clinico-pathological features of patients in Botswana with cervical dysplasia or cervical cancer

Table 2. Selected characteristics of patients in Botswana with cervical dysplasia vs cervical cancer

Figure Legends

Figure 1 Cervical microbiota of cervical dysplasia and cervical cancer in patients with and without HIV. A) Overall alpha diversity, as assessed by Shannon diversity in HIV-positive and -negative cervical dysplasia and cervical cancer patients. B) Beta diversity, as assessed by Bray-Curtis unweighted UniFrac in HIV-positive vs -negative patients. C) Stacked bar plot of the top 10 most abundant genus-level bacteria in HIV-positive vs -negative patients. Each bar represents a single patient and is labeled with the subject's age. D,E) LEfSe identified the most differentially abundant taxa between HIV-positive and -negative patients. D) Cladogram representation of the significantly different taxa features, from phylum (inner circle) to genus (outer circle). E) Histogram showing the LDA scores of genera that were differentially abundant between the 2 groups. The LEfSe was restricted to $p<0.05$ for the class and subclass analysis and a minimum LDA score of 2.0.

Figure 2 Cervical microbiota in cervical cancer patients is statistically significantly different from that in cervical dysplasia patients. A,B) Overall alpha diversity, as assessed by Shannon diversity in cervical dysplasia and cervical cancer patients. C,D) Beta diversity, as assessed by Bray-Curtis weighted UniFrac in cervical dysplasia vs cervical cancer patients. E) Stacked bar plot of the top 10 most abundant genus-level bacteria in cervical dysplasia patients vs cervical cancer patients. Each bar represents a single participant and is labeled with the subject's age. D,E) LEfSe identified the most differentially abundant taxa in cervical dysplasia and cervical cancer patients. D) Cladogram representation of the significantly different taxa features, from phylum (inner circle) to genus (outer circle). E) Histogram showing the LDA

scores of genera that were differentially abundant between the 2 groups. LEfSe was restricted to $p<0.05$ for the class and subclass analysis and a minimum LDA score of 2.0.

Table 1 Clinico-pathological features of 31 patients in Botswana with cervical dysplasia or cervical cancer

Feature	Result
Type of cervical lesion, n (%)	
CIN ~~stage~~ 1~~I~~	0
CIN ~~stage~~ 2~~II~~	3
CIN ~~stage~~ 3~~III~~	18
Cervical cancer	10
HIV status, %	
Positive	77
Negative	23
Smoking status, %	
Smoker	7
Non-Smoker	94

CIN, cervical intraepithelial neoplasia.

Commented [MOU1]: Some are numbers and some are % - clarify by putting n (5) for each variable. Make the % add up to 100%

Table 2 Selected characteristics of 31 patients in Botswana with cervical dysplasia vs. cervical cancer

Characteristic	Dysplasia (n=21)	Cancer (n=10)	P value*
Mean age (SD), years	41.8 (7.5)	50.7 (12)	0.1
Mean BMI (SD), kg/m^2	26 3 (6.4)	30.0 (7.2)	0.2
HIV status, %			
Positive	X (81%)	7 (70%)	0.5
Negative	X (19%)	3 (30%)	~~0.5~~
Smoking status, %			
Smoker	X (10%)	0 (0%)	0.3
Non-Smoker	X (91%)	10 (100%)	~~0.3~~

Commented [MOU1]: Add decimal point or round to add up to 100%

*P values were based on a t-test (continuous variables) or z-test (proportions). All tests were 2-sided.

Date : 4/11/2020 9:18:48 PM
From : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]
Cc : "Wong, Matthew C." [redacted]**, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org**
Subject : Re: MetaPhlan2

Hi Kristi,

Hope you are staying safe and healthy.

Wondering whether you have any update on the metaphlan2?

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Friday, March 27, 2020 10:52 AM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted] Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>; Petrosino, [redacted]
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud (and team),

The earliest the MetaPhlAn2 request can be completed is the week of April 6th. Let me know if you'd still like us to process the data given that timeframe.

Please note that with regards to Virmap, data processing requests need to go through a project manager and completed according to our queue. While we can expedite requests, especially for *trusted*, long-term collaborators, proper procedures still need to be followed. Circumventing these procedures affects other valued CMMR collaborators and is not taken lightly. I expect this won't be an issue going forward and any requests will go through the proper channels.

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 3:44 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted] Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>

Subject: Re: MetaPhlan2

Hi Kristi,

I am actually hoping to get the output of MetaPhlan2 by this week but if you can get it done by next week that would be great too.

I already got the output of VirMap. So, no worry on this analysis.

Best

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Wednesday, March 25, 2020 2:47 PM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted] ; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. <[redacted] ; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted]>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run them through the MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 3/25/2020 3:44:08 PM
From : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]
Cc : "Wong, Matthew C." [redacted] **"Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org**
Subject : Re: MetaPhlan2

Hi Kristi,

I am actually hoping to get the output of MetaPhlan2 by this week but if you can get it done by next week that would be great too.

I already got the output of VirMap. So, no worry on this analysis.

Best

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Wednesday, March 25, 2020 2:47 PM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted]>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run them through the MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 3/25/2020 2:48:17 PM
From : "Hoffman, Kristi Louise" [redacted]
To : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org
Cc : "Wong, Matthew C." [redacted]**, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org**
Subject : RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted] Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run them through the MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 3/27/2020 11:23:03 AM
From : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org
To : "Hoffman, Kristi Louise" [redacted]
Cc : "Wong, Matthew C." [redacted] **"Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "Petrosino, Joseph"** [redacted]
Subject : Re: MetaPhlan2

Hi Kristi,

Thanks for your email.

Yes, week of April 6 works as well. I want both the count and relative abundance data of the taxa. Thank you for your continued support.

Regards,

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Friday, March 27, 2020 10:52 AM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. <[redacted]>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>; Petrosino, Joseph[redacted]
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud (and team),

The earliest the MetaPhlAn2 request can be completed is the week of April 6th. Let me know if you'd still like us to process the data given that timeframe.

Please note that with regards to Virmap, data processing requests need to go through a project manager and completed according to our queue. While we can expedite requests, especially for *trusted*, long-term collaborators, proper procedures still need to be followed. Circumventing these procedures affects other valued CMMR collaborators and is not taken lightly. I expect this won't be an issue going forward and any requests will go through the proper channels.

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 3:44 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am actually hoping to get the output of MetaPhlan2 by this week but if you can get it done by next week that would be great too.

I already got the output of VirMap. So, no worry on this analysis.

Best

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Wednesday, March 25, 2020 2:47 PM
To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted]>; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run them through the MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 3/27/2020 10:52:45 AM
From : "Hoffman, Kristi Louise" [redacted]
To : "Khan,Md Abdul Wadud" MKhan7@mdanderson.org
Cc : "Wong, Matthew C." [redacted]**, "Ajami,Nadim J" NAjami@mdanderson.org, "Wargo,Jennifer" JWargo@mdanderson.org, "Petrosino, Joseph"** [redacted]
Subject : RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud (and team),

The earliest the MetaPhlAn2 request can be completed is the week of April 6th. Let me know if you'd still like us to process the data given that timeframe.

Please note that with regards to Virmap, data processing requests need to go through a project manager and completed according to our queue. While we can expedite requests, especially for *trusted*, long-term collaborators, proper procedures still need to be followed. Circumventing these procedures affects other valued CMMR collaborators and is not taken lightly. I expect this won't be an issue going forward and any requests will go through the proper channels.

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 3:44 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am actually hoping to get the output of MetaPhlan2 by this week but if you can get it done by next week that would be great too.

I already got the output of VirMap. So, no worry on this analysis.

Best

Wadud

From: Hoffman, Kristi Louise [redacted]
Sent: Wednesday, March 25, 2020 2:47 PM

To: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Cc: Wong, Matthew C. <[redacted] ; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: RE: MetaPhlan2

WARNING: This email originated from outside of MD Anderson. Please validate the sender's email address before clicking on links or attachments as they may not be safe.

Hi Wadud,

I can add your MetaPhlAn2 request to the Bioinformatics queue, but our BiT group is currently overwhelmed with other tasks so this won't be a quick turnaround. Is there a date by when you need these outputs?

Additionally, I've tried to find the Virmap bioinformatics request in our tracking system but haven't had much luck. Can you provide any further details on this?

Thanks,

Kristi

From: Khan,Md Abdul Wadud <MKhan7@mdanderson.org>
Sent: Wednesday, March 25, 2020 2:05 PM
To: Hoffman, Kristi Louise [redacted]
Cc: Wong, Matthew C. [redacted]; Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: Re: MetaPhlan2

Hi Kristi,

I am following up with you regarding running the WGS data through metaphlan2 pipeline and wondering whether there is any update on this.

Thank you

Wadud

From: Khan,Md Abdul Wadud
Sent: Friday, March 20, 2020 1:55 PM
To: Kristi Louise Hoffman [redacted]
Cc: [redacted] Ajami,Nadim J <NAjami@mdanderson.org>; Wargo,Jennifer <JWargo@mdanderson.org>
Subject: MetaPhlan2

Hi Kristi,

Recently, I shared WGS data with your group for running them through VirMap pipeline. I am wondering whether you could also run them through the

MetaPhlan2 pipeline for obtaining both the relative and absolute abundances of taxa as output. Here is the link for the WGS data: https://mdacc.app.box.com/folder/102021496910

I really appreciate your help and please let me know if you have questions.

Regards,

Wadud

The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.
The information contained in this e-mail message may be privileged, confidential, and/or protected from disclosure. This e-mail message may contain protected health information (PHI); dissemination of PHI should comply with applicable federal and state laws. If you are not the intended recipient, or an authorized representative of the intended recipient, any further review, disclosure, use, dissemination, distribution, or copying of this message or any attachment (or the information contained therein) is strictly prohibited. If you think that you have received this e-mail message in error, please notify the sender by return e-mail and delete all references to it and its contents from your systems.

Date : 5/15/2020 5:08:38 PM
From : "Ajami,Nadim J" najami@mdanderson.org
To : "Petrosino, Joseph" [redacted] "Hoffman, Kristi Louise" [redacted], "Javornik Cregeen, Sara Joan" [redacted], "Wong, Matthew C." [redacted]
Subject : VirMAP run

Hi Joe and team,

Torben Sølbeck, an investigator from the University of Copenhagen in the Dept. of Food Science is interested in using VirMAP to characterize the virome in a couple of datasets. They have developed their own pipeline and have used FastViromeExplorer but they aren't happy with either. Since we don't have a solution available for external users (yet), I wanted to ask for your help with this. He has made the dataset available to download (18Gb compressed tarball) – should be an easy and quick run for Matt if you are interested in helping him out. Of course, anything that comes out of this will be properly referenced and acknowledged.

Here's the link:
https://filesender.deic.dk/?s=download&token=e8f04acd-5c13-f749-2d91-e2ca12e6d128

Hope you are all well,
Nadim